THE ANNALS
of
MATHEMATICAL STATISTICS

THE ANNALS OF MATHEMATICAL STATISTICS IS AFFILIATED WITH THE AMERICAN STATISTICAL ASSOCIATION AND IS DEVOTED TO THE THEORY AND APPLICATION OF MATHEMATICAL STATISTICS

Volume VI, 1935

PUBLISHED QUARTERLY
ANN ARBOR, MICHIGAN

Four Dollars per annum

Made in United States of America

Address: ANNALS OF MATHEMATICAL STATISTICS
Post Office Box 171, Ann Arbor, Michigan

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD.

CONTENTS OF VOLUME VI

CONTENTS

SOME INTERESTING FEATURES OF FREQUENCY CURVES

By Richmond T. Zoch

Introduction

It is well known that in the normal error curve the points of inflection are equidistant from the mode. However it has never been pointed out that this is also a characteristic of all of the bell-shaped Pearson Frequency Curves. This fact can be most easily shown by placing the mode at the abscissa $x = 0$.

Many rough checks have been developed for use in applying the Theory of Least Squares. The second part of this paper develops a rough check on the computation for use when fitting a Pearson Frequency Curve to a set of observations. No rough checks on computation are given in textbooks on Pearson's Frequency Curves.

At present it is customary to follow a separate procedure for each Type of curve when computing the constants of a Pearson Frequency Curve. The third part of this paper shows how a single system may be followed for all Types. A single procedure is very desirable in order that the rough check of Part 2 may be quickly applied.

Part 1. Points of Inflection

Perhaps nothing brings out the limitations of the bell-shaped Pearson Curves in a more striking manner than a discussion of their points of inflection. In dealing with frequency curves it is well known that any curve can be fitted to a given distribution and that the real problem in curve fitting is the selection of a curve. Figures 1, 2, and 3 illustrate three hypothetical histograms. All three of these histograms are bell-shaped yet none of them will be closely fitted by any of the Pearson Curves. The reasons will be pointed out presently.

The differential equation from which Pearson derived his system of frequency curves is

$$\frac{dy}{dx} = \frac{y(x - P)}{b_2x^2 + b_1x + b_0}.$$

By putting $x - P = X$, i.e. by placing the mode at the abscissa $X = 0$, this differential equation may be written:

$$\frac{dy}{dX} = \frac{yX}{\pm B_2X \pm B_1X + B_0}$$

where the $+$ or $-$ sign is taken according to the type of the curve. (It will be shown later that the constant term of the denominator must be less than zero.)

Since in the Type III curve B_2 is 0 and in the "Normal Curve" both B_2 and B_1 are 0 it will be advantageous to consider the general case of

$$\frac{dy}{dX} = \frac{yX}{F(X)},$$

where $F(X)$ is an integral rational function of the n^{th} degree, at once rather than considering special cases first.

If

$$\frac{dy}{dX} = \frac{yX}{F(X)},$$

then

$$\frac{d^2y}{dX^2} = \frac{y}{[F(X)]^2}\left\{X^2 + F(X) - XF'(X)\right\}.$$

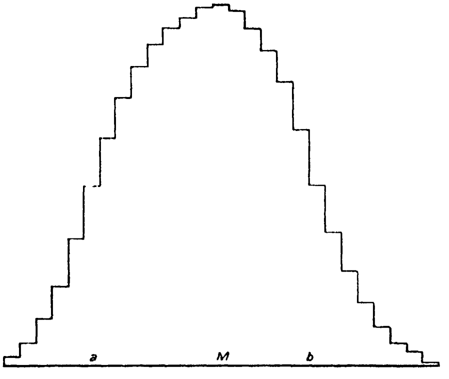


FIG. 1

In order to locate the points of inflection, $\frac{d^2y}{dX^2}$ is equated to zero. Then we have:

$$X^2 + F(X) - XF'(X) = 0. \tag{1}$$

This equation is always of the same degree as $F(X)$ except when $F(X)$ is linear or constant. Hence we have proved the Theorem: If $y = G(X)$ be the solution of the differential equation

$$\frac{dy}{dX} = \frac{yX}{F(X)},$$

then the number of points of inflection of y cannot exceed the degree of $F(X)$ when $F(X)$ is of degree greater than one.

Now $F(X) = B_nX^n + B_{n-1}X^{n-1} + \cdots + B_2X^2 + B_1X + B_0$. Whence equation (1) can be written in the form:

$$(1 - n)B_nX^n + (2-n)B_{n-1}X^{n-1} + (3 - n)B_{n-2}X^{n-2} + \cdots$$
$$+ (r - n)B_{n-r+1}X^{n-r+1} + \cdots - 3B_4X^4 - 2B_3X^3 + (1 - B_2)X^2 + B_0 = 0.$$

Hence we have established the Theorem: The coefficient of the linear term of X in the equation of the points of inflection is zero.

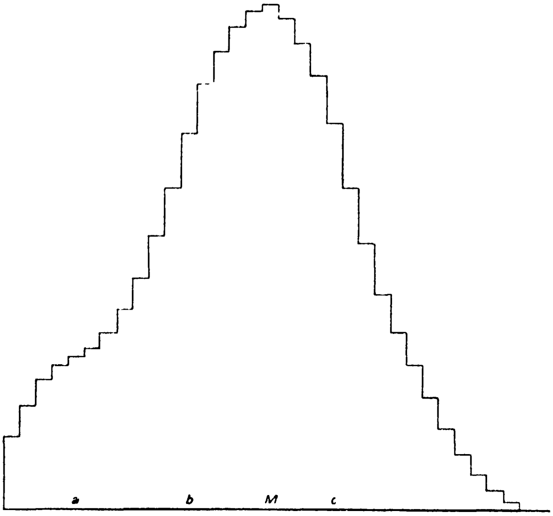


FIG. 2

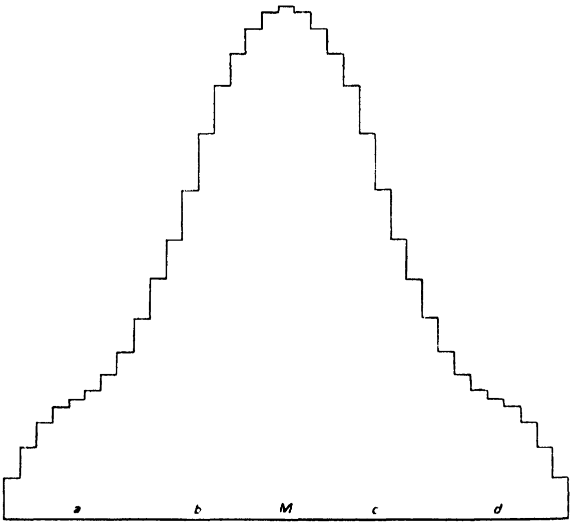


FIG. 3

For the "Normal Curve" and also for Type III,

$$B_2 = B_3 = B_4 = \cdots = B_n = 0.$$

Hence the points of inflection of these two Types are given by $X = \pm\sqrt{-B_0}$.

For Types I and II, B_2 is positive and $B_3 = B_4 = \cdots = B_n = 0$, and the

points of inflection are $X = \pm\sqrt{\dfrac{-B_0}{1 - B_2}}$. Hence the points of inflection are undefined if $B_2 = 1$, are pure imaginary if $B_2 > 1$, and real if $B_2 < 1$.

For Types IV, V, VI and VII, B_2 is negative and $B_3 = \cdots = B_n = 0$, and the points of inflection are at $X = \pm\sqrt{\dfrac{-B_0}{1 + |B_2|}}$.

In some of these Types it may happen that the abscissae of the points of inflection though real will lie beyond the range of the curve. Thus Types III and VI may have 1 or 2 points of inflection, the single point of inflection occurring when $\sqrt{\dfrac{-B_0}{1 + B_2}}$ > the range of the curve in the direction that the range is limited. Type II may have 0 or 2 points of inflection, as there will be no real points of inflection when $B_2 \geqq 1$. Type I may have 0, 1 or 2 points of inflection. Types IV, V and VII as well as the "Normal Curve" always have 2 and only two points of inflection.

Now it should be noted that when one of the eight bell-shaped Pearson curves has two points of inflection then the abscissae of these 2 points of inflection are equidistant from the abscissa of the mode. In figure 1 a point of inflection will be at abscissa b and another at abscissa a. (M is the abscissa of the mode.) Since $b - M \neq M - a$ none of the Pearson curves will fit this histogram closely. In figure 2, points of inflection occur at abscissae a, b, and c. Since a Pearson curve can have at most two points of inflection no Pearson curve will fit this histogram closely. In figure 3 there are four points of inflection and no Pearson curve will fit this histogram closely.

Part 2. Range

Definition: A bell-shaped curve is a continuous curve which starts at zero (or zero as a limit), rises to a single maximum, at which maximum point the first derivative is zero, and then falls to zero (or zero as a limit).

Or, more formally, $y = G(x)$ is a bell-shaped curve if $G(x_1) = G(x_2) = 0$ and if $G'(P) = 0$ and $G''(P) < 0$ where $G(x)$ is continuous and does not vanish in the interval from x_1 to x_2 and P is a unique point in this interval.

If a bell-shaped curve has the value of zero at two finite points, one on each side of the maximum (mode), it is said to be of limited range in both directions, or briefly, of limited range.

If a bell-shaped curve has the value of zero at only one finite point it is said to be of limited range in one direction, or also of unlimited range in one direction.

If a bell-shaped curve has the value of zero only at $\pm\infty$, i.e. at no finite points, it is said to be of unlimited range in both directions, or briefly, of unlimited range.

Theorem I: If $F(x)$ can be separated into a finite number of factors each either of the form $(x - r_i)$ or $(x^2 + 2r_j\, x + r_j^2 + r_{0j}^2)$ where no real root is repeated and $y = G(x)$ is a *bell-shaped* curve which is a solution of the differential equation

$$\frac{dy}{dx} = \frac{y(x - P)}{F(x)},$$

then if $F(x)$ has no real roots, y is of unlimited range in both directions; if all of the real roots of $F(x)$ lie on the same side of P, y is of limited range in one (that) direction; if at least one real root of $F(x)$ lies on one side of P and at least one on the other side, y is of limited range in both directions.

Proof: If $F(x) = 0$ when $x = P$, we have

$$\frac{dy}{dx} = \frac{y}{g(x)}$$

where $g(x) = F(x) \div (x - P)$. This derivative is zero only when $y = 0$ or $g(x) = \pm\infty$. Hence the solution does not have a finite maximum and therefore is not a bell-shaped curve. If $F(x) > 0$ when $x = P$, we have

$$\frac{d^2y}{dx^2} = \frac{y}{[F(x)]^2}\left[(x - P)^2 + F(x) - (x - P)\frac{d}{dx}F(x)\right]$$

$$\frac{d^2y}{dx^2} \qquad = \frac{y}{[F(x)]^2}[F(x)]$$

which is greater than zero and, since at a maximum the second derivative must not be greater than zero, in this case the solution would have a minimum at $x = P$ and therefore would not be a bell-shaped curve. As the theorem concerns only those solutions which are bell-shaped curves, $F(x) < 0$ when $x = P$. If $F(x) = 0$ when $x \neq P$ then $\frac{dy}{dx} = \pm\infty$ unless y is also zero. Assume $y \neq 0$. Since $F(x)$ is negative, if $y \neq 0$ when $F(x) = 0$ then $\frac{dy}{dx} \to -\infty$ as $F(x) \to 0$, for an $x > P$, and changes to $+\infty$ as $F(x)$ changes sign on passing through the value 0. Hence the curve would contain another maximum before falling to zero and therefore the solution is not a bell-shaped curve. Similar reasoning holds for an $x < P$. Therefore if $y \neq 0$ when $F(x) = 0$, the curve is not bell-shaped. If $y = 0$ when $F(x) = 0$, the curve has its range limited at this point. That is, any real number which makes $F(x)$ vanish will also make y vanish if y represents a bell-shaped curve. Hence if all of the real roots lie on the same side of P the curve is of limited range in that direction only, while if at least one of the real roots lies on each side of P the curve is of limited range in both directions. If $F(x)$ contains no real roots it does not vanish for any real value of x. In this case, by partial fractions the differential equation becomes:

$$\frac{dy}{y} = \frac{k_{11}\,dx}{(x + r_1)^2 + r_{01}^2} + \frac{k_{21}\,dx}{(x + r_2)^2 + r_{02}^2} + \cdots + \frac{2k_{21}(x + r_1)\,dx}{(x + r_1)^2 + r_{01}^2}$$

$$+ \frac{2k_{22}(x + r_2)\,dx}{(x + r_2)^2 + r_{02}^2} + \cdots.$$

On integrating,

$$y = C\left[(x + r_1)^2 + r_{0_1}\right]^{k_{21}}\left[(x + r_2)^2 + r_{0_2}^2\right]^{k_{22}} \cdots e^{k_{11} \arctan \frac{\quad}{r_{0_1}}}$$

Hence y does not vanish for a finite real value of x and the Theorem is fully established.

THEOREM II: If $F(x)$ can be separated into a finite number of factors each either of the from $(x - r_i)$ or $(x^2 + 2r_jx + r_j^2 + r_{0_j}^2)$ where no real root is repeated and $y = G(x)$ is a *bell-shaped* curve which is a solution of the differential equation $\frac{dy}{dx} = \frac{y(x - P)}{F(x)}$, then if y is of unlimited range, $F(x)$ contains no real roots; if y is of limited range in one direction, all of the real roots of $F(x)$ lie on the same (that) side of P; if y is of limited range in both directions, at least one of the real roots of $F(x)$ lies on one side of P and at least one on the other.

PROOF: By partial fractions the differential equation may be written:

$$\frac{k_{11}\,dx}{x - r_{11}} + \frac{k_{21}\,dx}{x - r_{12}} + \cdots + \frac{k_{21}\,dx}{(x + r_{21})^2 + r_{0_1}}$$

$$+ \frac{k_{22}\,dx}{(x + r_{22})^2 + r_{0_2}^2} + \cdots \frac{2k_{31}(x + r_{21})\,dx}{(x + r_{21})^2 + r_{0_1}^2} + \frac{2k_{32}(x + r_{23})\,dx}{(x + r_{22})^2 + r_{0_2}^2} +$$

and on integrating:

$$y = C(x - r_{11})^{k_{11}}(x - r_{12})^{k_{12}} \cdots \left[(x + r_{21})^2 + r_{0_1}^2\right]^{k_{31}} \cdots e^{k_{21} \arctan \frac{\quad}{r_{0_1}} + \cdots}.$$

Hence $y = 0$ for $x = r_{11}, r_{12}, \cdots$ and for no other finite values of x provided k_{11}, $k_{12}, \cdots$ are positive. If one or more of the k_{ij} are negative, $y = \infty$ at such points and unless some r_{ij} closer to P has previously made y vanish, the curve is not bell-shaped. Therefore, for bell-shaped curves, the exponent of the factor containing the real root of smallest absolute value on each side of P is positive. Therefore: if y is of limited range in both directions, at least one real root lies on each side of P; if y is of unlimited range in one direction, all of the real roots lie on the same side of P; if y is of unlimited range it contains no real roots. Hence the Theorem is established.

The effect of repeated real roots will now be considered. If a real root is repeated an odd number of times at $x = r$, then $F(x)$ changes sign at $x = r$ and the first theorem is true. If a real root is repeated an even number of times at $x = r$, then $F(x)$ does not change sign at $x = r$ and we know that either (a) $y = 0$ at $x = r$; or (b) y is finite and $\neq 0$ and $\frac{dy}{dx} = \pm\infty$ at $x = r$, i.e. there is a point of inflection at $x = r$. It will now be shown that (b) cannot occur. If case (b) is possible, y is continuous at $x = r$, $\frac{dy}{dx} = \pm\infty$ according as $(r - P) \lessgtr 0$

moreover $\frac{dy}{dx}$ does not change sign in the neighborhood of the point $x = r$, and $\frac{d^2y}{dx^2}$ changes sign from $+\infty$ to $-\infty$ or vice versa according as $(r - P) \lessgtr 0$. Now

$$\frac{d^2y}{dx^2} = \frac{y}{[F(x)]^2}\left[(x - P)^2 + F(x) - (x - P)\frac{d}{dx}F(x)\right].$$

Whence if y is finite and $\neq 0$, $\frac{d^2y}{dx^2}$ does not change sign at $x = r$ because it is possible to select a neighborhood such that

$$(x - P)^2 \mid > F(x) - (x - P)\frac{d}{dx}F(x)$$

for an x differing from r by ϵ where ϵ is a small positive quantity. Therefore case (b) is not possible and $y = 0$ when a real root is repeated an even number of times. That is to say the range of the curve is limited at a point where a real root is repeated an even number of times. Thus Theorem I always holds for repeated roots.

For Theorem II it is clear that this Theorem holds for repeated roots when a non-repeated root lies closer to P, and on the same side, than the repeated root. Suppose that the repeated root is the nearest root to P (on a given side of P). Then by partial fractions:

$$\frac{dy}{y} = \frac{k_{11}\,dx}{(x - r_{11})} + \frac{k_{12}\,dx}{(x - r_{11})^2} + \frac{k_{13}\,dx}{(x - r_{11})^3} + \cdots + \frac{k_{41}\,dx}{(x - r_{41})} + \frac{k_{42}\,dx}{(x - r_{42})}$$
$$+ \cdots + \frac{k_{21}\,dx}{(x + r_{21})^2 + r_{0_1}^2} + \frac{k_{22}\,dx}{(x + r_{22})^2 + r_{0_2}^2} + \cdots + \frac{2k_{31}(x + r_{21})\,dx}{(x + r_{21})^2 + r_{0_1}^2} +$$

and on integrating:

$$y = C(x - r_{11})^{k_{11}}(x - r_{41})^{k_{41}}(x - r_{42})^{k_{42}} \cdots \left[(x + r_{21})^2 + r_{0_1}^2\right]^{k_{31}}$$
$$\cdots e^{k_{21}\,\text{arc tan}\,\frac{x+r}{r_{0_1}} \quad (x - r_{11}) \quad 2(x - r_{11})^2}.$$

Hence y can $= 0$ only for $x = r_{11}$ or for $x = r_{41}, r_{42}, \cdots$ and for no other finite values of x. Since by hypothesis y is bell-shaped, then the proper k_{ij} must be positive and Theorem II always holds for repeated roots.

Theorems I and II can now be combined and generalized in the form:

Theorem: If $F(x)$ is a polynomial with real coefficients and $y = G(x)$ is a bell-shaped curve which is a solution of the differential equation

$$\frac{dy}{dx} = \frac{y(x - P)}{F(x)},$$

then the necessary and sufficient condition: that y be of unlimited range in both directions is that $F(x)$ have no real roots; that y be of limited range in one direction is that all of the real roots of $F(x)$ lie on the same side of P; that y be of limited range in both directions is that at least one real root of $F(x)$ lie on one side of P and one on the other.

COROLLARY: $F(x)$ must be negative throughout the range of y.

Suppose now that we have some statistics which we wish to graduate and the statistics are of such nature that we would expect a bell-shaped curve, rather than a J- or U-shaped curve, and we desire the best fit: If we use a curve which is a solution of the differential equation

$$\frac{dy}{dx} = \frac{y(x - P)}{F(x)}$$

(the Pearson Curves being special cases) to fit the statistics and if in computing the constants for the curve one of the following cases arise:

(a) $b_0' \nless 0$ when this constant is computed,

or (b) $B_0 \nless 0$ when the origin is moved to the mode,

or (c) a root is located within the range of the statistics then it means that:

1. A mistake may have been made in the computation: thus the Theorem just established provides a rough check on the work of computation,

2. If no mistake has been made in the computation it may indicate that the bell-shaped Pearson Curves will not closely fit the statistics and that some other graduation curves be used, e.g. the Gram-Charlier Types A or B might be tried,

3. If no mistake has been made in the computation it may happen that one of the bell-shaped Pearson Curves will give an excellent fit but a different method than or a modification of the Method or Moments should be used in order to compute the constants.

Part 3. Computing the Constants

At present, the constants of a frequency curve are computed as follows: First the moments are computed about an arbitrary origin, then the moments about the A.M. are determined, then β_1 and β_2 and the criterion are computed, after which the type of curve can be selected. From this point a separate procedure is followed for each curve. Now in the above method one will not know whether a root has been located in the range of statistics or not.

Take Pearson's differential equation

$$\frac{dy}{dx} = \frac{y(x - P)}{b_2x^2 + b_1x + b_0}.$$

Put $X = x - P$. Then $dX = dx$ and $x = X + P$, and

$$\frac{dy}{dx} = \frac{yX}{b_2(X + P)^2 + b_1(X + P) + b_0} = \frac{yX}{b_2X^2 + 2Pb_2X + b_1X + P^2b_2 + Pb_1 + b_0}.$$

Now put

$$b_2 = B_2$$
$$2Pb_2 + b_1 = B_1$$
$$P^2b_2 + Pb_1 + b_0 = B_0\,.$$

Then we have

$$\frac{dy}{dX} = \frac{yX}{B_2X^2 + B_1X + B_0} \quad \text{or} \quad \frac{dy}{dx} = \frac{y(x-P)}{B_2(x-P)^2 + B_1(x-P) + B_0}\,. \tag{1}$$

It should be noted that for a particular curve, B_2, B_1 and B_0 are constants; i.e., their values do not change with a change of the origin. The values of b_1 and b_0 do change with a change in the origin.

If we clear equation (1) of fractions, multiply by $e^{\eta x}$ and integrate with respect to x over the range from x_1 to x_2, where

$$e^{\lambda_1\eta + \frac{\lambda_2\eta^2}{2!} + \frac{\lambda_3\eta^3}{3!} + \cdots} \equiv \int_{x_1}^{x_2} e^{\eta x}\, y dx\,,$$

then successively differentiate with respect to η, and equate coefficients of like powers of η, we finally obtain:

$$\left.\begin{aligned}
&\lambda_1 - P + B_1 - 2PB_2 + 2B_2\lambda_1 = 0\,,\\
&\lambda_2 + B_0 - PB_1 + P^2B_2 + B_1\lambda_1 - 2PB_2\lambda_1 + 3B_2\lambda_2 + B_2\lambda_1^2 = 0\,,\\
&\lambda_3 + 2\lambda_2B_1 - 4PB_2\lambda_2 + 4B_2\lambda_3 + 4B_2\lambda_1\lambda_2 = 0\,,\\
&\lambda_4 + 3B_1\lambda_3 - 6PB_2\lambda_3 + 5B_2\lambda_4 + 6B_2\lambda_2^2 + 6B_2\lambda_1\lambda_3 = 0\,.
\end{aligned}\right. \tag{2}$$

Since we can compute the moments from the raw statistics and the semi-invariants from the moments, we may regard λ_2, λ_3 and λ_4 in these equations as knowns and the B_0, B_1, B_2, P and λ_1 as unknowns. But the origin has not yet been specified. Let the origin be placed at the A.M. where $\mu_1 = \lambda_1 = 0$. As λ_2, λ_3, λ_4, B_0, B_1 and B_2 are unchanged by a change of origin, we have:

$$\left.\begin{aligned}
&B_1 - P_0 - 2P_0B_2 = 0\,.\\
&\lambda_2 + B_0 - P_0B_1 + P_0^2B_2 + 3B_2\lambda_2 = 0\,,\\
&\lambda_3 + 2B_1\lambda_2 - 4P_0B_2\lambda_2 + 4B_2\lambda_3 = 0\,,\\
&\lambda_4 + 3B_1\lambda_3 - 6P_0B_2\lambda_3 + 5B_2\lambda_4 + 6B_2\lambda_2^2 = 0\,.
\end{aligned}\right. \tag{3}$$

Now put

$$\left.\begin{aligned}
b_0' &= B_0 - P_0B_1 + P_0^2B_2\,,\\
b_1' &= B_1 - 2P_0B_2\,,\\
b_2' &= B_2\,;
\end{aligned}\right\} \tag{4}$$

then

$$
\begin{aligned}
&b_1' - P_0 = 0\,, \\
&\lambda_2 + b_0' + 3b_2'\lambda_2 = 0\,, \\
&\lambda_3 + 2b_1'\lambda_2 + 4b_2'\lambda_3 = 0\,, \\
&\lambda_4 + 3b_1'\lambda_3 + 5b_2'\lambda_4 + 6b_2'\lambda_2^2 = 0\,.
\end{aligned}
\tag{5}
$$

By reversing the transformation (4) we get:

$$
\left.
\begin{aligned}
B_2 &= b_2'\,, \\
B_1 &= b_1' + 2P_0b_2' \\
B_0 &= b_0' + P_0(b_1' + P_0b_2')\,.
\end{aligned}
\right\}
\tag{6}
$$

Now the above theory suggests the following procedure for computing the constants of a frequency curve: First the moments are computed about an arbitrary origin, then the semi-invariants are computed (or alternatively the moments about the A.M., either step involves about the same amount of work), then the equations (5) are solved and then by means of equations (6) the B_2, B_1 and B_0 are computed. Next solve the quadratic equation

$$B_2X^2 + B_1X + B_0 = 0.$$

The character of the roots of this equation indicates which type to use and it is unnecessary to compute the criterion. The constants of the frequency curve are simple functions of the roots of the above quadratic equation and can be readily found by integrating the diff. eq. (1) being careful to write the solution as a function of $X = x - P$. The rough checks mentioned in Part 2 can be quickly and conveniently applied when this procedure is followed.

GEORGE WASHINGTON UNIVERSITY.

A RECONSIDERATION OF SHEPPARD'S CORRECTIONS

By W. T. Lewis[1]

In computing the moments of a frequency distribution it is customary to find first what are known as the raw moments. These are obtained on the assumption that all the material of each class interval is concentrated at the middle point of the interval. It introduces what is called a grouping error because in fact the material does not all lie at the middle point. To compensate for this error W. F. Sheppard[2] derived a set of corrections. The hypothesis underlying his method is that the distribution may be regarded as similar to one to which the Euler-MacLaurin summation formula without its end terms may be applied. He presupposed such a curve, found its true moments, and then the raw moments that would be obtained if its area were concentrated at several equidistant abscissae. The relationship between these raw moments and the true moments of the curve furnished him with the corrections required for that distribution. If now our observed distribution may be supposed to be sufficiently like that one, we may use his corrections also on the observed data. One may note four points of criticism.

(1) The given distribution may not be similar to the one suggested, in the sense that it would be close to such a curve if the intervals of grouping were made very small; or at all events the purpose of finding the moments may be in part to decide whether or not it would become such a curve, and so one would not like to assume that to be true at the outset. A special case of importance in which this last is true occurs when one is finding the moments of a sample in order to determine whether it may have been drawn from a presupposed universe. It is inexact to use raw moments but it is illogical to use corrections that have been proved only for the universe being tested.

(2) Sheppard's argument does not make use of the one certain fact that is given in the hypothesis, viz: that the partial area of the given distribution over each class interval is exactly as stated. In fact, if, following the argument of some authors, the given curve be assumed to be exponential, it obviously cannot have partial areas everywhere exactly equal to the several given frequencies, for in particular its partial area is not zero beyond the given range.

(3) It is common to find distributions which do not have high contact at the ends of the range and for them Sheppard's corrections certainly fail. To obviate this criticism new corrections have been derived by Pairman and Pear-

[1] With the assistance of Burton H. Camp.

[2] The true values are given on page 220 of "Mathematical Part of Elementary Statistics, by Camp, D. C. Heath and Company, 1931.

son for the so-called abrupt cases. These new corrections are adequate to care for the abrupt cases but involve so much computation that it is a fair question whether it would not be simpler, first to distribute the given material over each interval by a smoothing process, and then to find without corrections the moments of the smoothed distribution.

(4) Even if one admits Sheppard's method in general, waiving the dubious question as to whether it is proper to start with an assumed curve instead of starting with the given distribution, it is doubtful whether there are any curves which have exactly the properties required. The high contact hypothesis may be put in different language as follows: using the notation of the Handbook[3] page 92, let $f(x)$ be the curve and x_t be the middle point of the slice. It is assumed that

$$\sum_t c x_t^r f^{(i)}(x_t) = \int_{-\infty}^{\infty} x^r f^{(i)}(x)\, dx\,; \qquad i = 0, 1, \cdots\,; \qquad r = 0, 1, \cdots\,;$$

c being the class interval. This means that if the moments of the curve be found by using *mid-ordinates times class interval*, instead of *areas*, one will obtain exactly the true moments of the curve, and that this will remain true for all the curves which are derivatives of this curve. This property is certainly not true of the normal curve; but it is almost true when r and the class interval are both small, and it is probably due to this fact that Sheppard's corrections seem to be good in practice.

Moreover, this high contact hypothesis cannot be true for any function over a limited range if the function is developable in Taylor's series about one end of the range. For the only function which has the required properties is identically zero, since the function and all its derivatives are required to vanish at that end of the range.

The primary purpose of this paper, therefore is to derive corrections similar to Sheppard's with a different set of assumptions. The results may be used as an approximate substitute for both Sheppard's and Pairman's. That is, they will apply approximately to both extreme cases and to the intermediate cases; on the whole they give better results than Sheppard's and are not so difficult to administer as Pairman's.

The argument runs as follows. When a distribution is given merely by class intervals, there is no way of knowing exactly what the distribution would have been had the class intervals been smaller; we do not know that we have a sample from an exponential curve, and even if we did we would not know that this sample would lie close to the exponential in form. We shall, however, try to draw a graduating curve in such a manner that (a) its partial area over each class interval will equal the frequency of the given distribution over that interval; and (b) its form within each class interval will be such that it will pass smoothly into the adjacent portions to the right and left. A good way to do this is by a

[3] H. L. Rietz, "Handbook of Math. Stat." Houghton Mifflin Co. (1924).

freehand graph, frankly recognizing that there are many forms that will do equally well. To obtain a numerical result it is necessary to use the equation of some curve. Again frankly recognizing that there are many types which will do equally well we choose the simplest to handle:

$$y = a + bt + ct^2.$$

Let the relative frequency distribution be defined by $f(i)$, $-m \leqq i \leqq n$, m, n, i being integers. To satisfy (a) we have the equation

$$\int_{i-\frac{1}{2}}^{i+\frac{1}{2}} y\,dt = f(i).$$

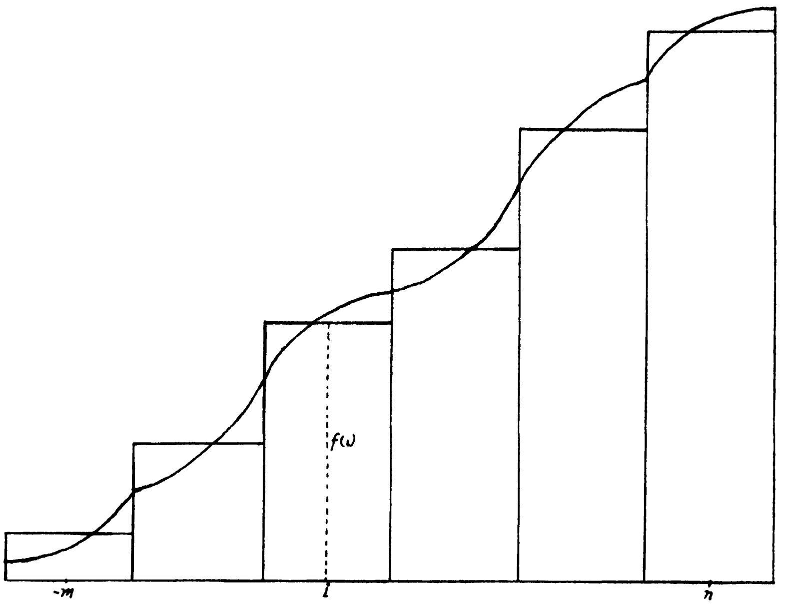


FIG. 1

To satisfy (b) we shall let

$$y = \tfrac{1}{2}[f(i) + f(i + 1)] \text{ if } t = i + \tfrac{1}{2}.$$

The latter will hold for all values of i from $-m$ to $n - 1$ inclusive, but the end intervals require special treatment. Here in order to satisfy as well as possible both the high contact and the abrupt cases, we wish to let the material be distributed according to the way the curve is behaving over the two nearest intervals on the right (at n) or left (at $-m$) rather than by the addition of zero frequencies beyond the given limits. To do this we let the slope of the parabolas be zero at the extremes:

$$\frac{dy}{dt} = 0 \qquad \text{at } t = -m - \tfrac{1}{2} \text{ and } t = n + \tfrac{1}{2}.$$

Then, if for example the frequencies are increasing as one nears the right end interval, the curve will rise over the right end interval; if they are decreasing, it will fall. These three conditions are sufficient to determine a continuous curve of the sort indicated in the figure. The exact moments of the curve may be found by integration and expressed in terms of the raw moments. The details are tedious and of an elementary nature and will be given only for the mean value $\bar{\nu}_1$.

To determine the coefficients of the parabola $y = a + bt + ct^2$ for the rectangle at $t = i$ we may write the following three equations; the first complying with the requirement that the area under the parabola from $t = i - \frac{1}{2}$ to $t = i + \frac{1}{2}$ equals the area of the rectangle at $t = i$, the second and third giving the ordinates at $i - \frac{1}{2}$ and $i + \frac{1}{2}$ respectively:

$$f(i) = \int_{i-\frac{1}{2}}^{i+\frac{1}{2}} (a + bt + ct^2)dt\,,$$

$$\frac{f(i) + f(i+1)}{2} = a + b\,(i + \tfrac{1}{2}) + c\,(i + \tfrac{1}{2})^2\,,$$

$$\frac{f(i) + f(i-1)}{2} = a + b\,(i - \tfrac{1}{2}) + c\,(i - \tfrac{1}{2})^2\,.$$

Solving these three simultaneous equations we get for a, b, and c:

$$a = (\tfrac{5}{4} - 3i^2)\,f(i) + \left(\frac{3i^2}{2} - \frac{i}{2} - \frac{1}{8}\right)f(i+1) + \left(\frac{3i^2}{2} + \frac{i}{2} - \frac{1}{8}\right)f(i-1)\,,$$

$$b = 6if(i) + (\tfrac{1}{2} - 3i)f(i+1) - (\tfrac{1}{2} + 3i)f(i-1)\,,$$

$$c = -\,3f(i) + \tfrac{3}{2}\,f(i+1) + \tfrac{3}{2}\,f(i-1)\,,$$

and these hold for $-m + 1 \leqq i \leqq n - 1$.

For the parabola $y = a_1 + b_1t + c_1t^2$ over the first rectangle, i.e., where $i = -m$, we get the equations:

$$f(-m) = \int_{-m-\frac{1}{2}}^{-m+\frac{1}{2}} (a_1 + b_1\,t + c_1\,t^2)dt\,,$$

$$\frac{f(-m) + f(-m+1)}{2} = a_1 + b_1\,(-m + \tfrac{1}{2}) + c_1\,(-m + \tfrac{1}{2})^2\,,$$

$$b_1 + 2\,c_1\,(-m - \tfrac{1}{2}) = 0\,,$$

and their solutions:

$$a_1 = \tfrac{3}{4}\,(m^2 + m - \tfrac{1}{12})\,f(-m+1) - \tfrac{3}{4}\,(m^2 + m - \tfrac{17}{12})\,f(-m)\,,$$

$$b_1 = \tfrac{3}{4}\,(2m+1)\,f(-m+1) - \tfrac{3}{4}\,(2m+1)\,f(-m)\,,$$

$$c_1 = \tfrac{3}{4}\,f(-m+1) - \tfrac{3}{4}\,f(-m)\,.$$

Similarly for the parabola $y = a_n + b_n t + c_n t^2$ through the last rectangle at $i = n$ we get

$$f(n) = \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} (a_n + b_n t + c_n t^2)dt\,,$$

$$\frac{f(n) + f(n-1)}{2} = a_n + b_n\,(n - \tfrac{1}{2}) + c_n\,(n - \tfrac{1}{2})^2\,,$$

$$b_n + 2\,c_n n + c_n = 0\,,$$

and for the constants

$$a_n = \tfrac{3}{4}\,(n^2 + n - \tfrac{1}{12})\,f(n-1) - \tfrac{3}{4}\,(n^2 + n - \tfrac{17}{12})\,f(n)\,,$$

$$b_n = -\,\tfrac{3}{4}\,(1 + 2n)\,f(n-1) + \tfrac{3}{4}\,(1 + 2n)\,f(n)\,,$$

$$c_n = \tfrac{3}{4}\,f(n-1) - \tfrac{3}{4}\,f(n)\,.$$

Having obtained the constants for the graduating curve we will determine the moments of this curve in terms of those of the given frequency distribution.

Notation: Let the class interval be $c = 1$; let $\nu_s = \sum_{i=-m}^{n} i^s f(i)$ be the uncorrected s^{th} moment of the given frequency distribution about the given origin; let $\mu_s = \sum_{i=-m}^{n} (i - \nu_1)^s f(i)$ be the uncorrected s^{th} moment of the given frequency distribution about its uncorrected mean; let $\bar{\nu}_s$ be the corrected value of the s^{th} moment about the given origin; and let $\bar{\mu}_s$ be the corrected value of the s^{th} moment about the corrected mean. Thus ν_s and μ_s apply to the rectangles, and $\bar{\nu}_s$ and $\bar{\mu}_s$ apply to the curves as follows:

$$\bar{\nu}_s = \sum_{i=-m+1}^{n-1} \int_{i-\frac{1}{2}}^{i+\frac{1}{2}} t^s(a + bt + ct^2)dt + \int_{-m-\frac{1}{2}}^{-m+\frac{1}{2}} t^s(a_1 + b_1 t + c_1 t^2)dt$$
$$+ \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} t^s(a_n + b_n t + c_n t^2)dt\,,$$

$$\bar{\mu}_s = \sum_{i=-m+1}^{n-1} \int_{i-\frac{1}{2}}^{i+\frac{1}{2}} (t - \bar{\nu}_1)^s\,(a + bt + ct^2)dt + \int_{-m-\frac{1}{2}}^{-m+\frac{1}{2}} (t - \bar{\nu}_1)^s\,(a_1 + b_1 t + c_1 t^2)dt$$
$$+ \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} (t - \bar{\nu}_1)^s\,(a_n + b_n t + c_n t^2)dt\,.$$

Using these symbols we have for the first moment about the given origin:

$$\bar{\nu}_1 = \sum_{i=-m+1}^{n-1} \int_{i-\frac{1}{2}}^{i+\frac{1}{2}} t(a + bt + ct^2)dt + \int_{-m-\frac{1}{2}}^{-m+\frac{1}{2}} t(a_1 + b_1 t + c_1 t^2)dt$$
$$+ \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} t(a_n + b_n t + c_n t^2)dt$$

$$= \sum_{-m+1}^{n-1}\left[ai + b\left(i^2 + \frac{1}{12}\right) + c\left(i^3 + \frac{i}{4}\right)\right]$$

$$+\left[-a_1 m + b_1\left(m^2 + \frac{1}{12}\right) - c_1\left(m^3 + \frac{m}{4}\right)\right]$$

$$+\left[a_n n + b_n\left(n^2 + \frac{1}{12}\right) + c_n\left(n^3 + \frac{n}{4}\right)\right].$$

Substituting the values for the constants this becomes

$$\bar{\nu}_1 = \sum^{n-1} \Big\{ i\left[\left(\frac{5}{4} - 3i^2\right)f(i) + \left(\frac{3i^2}{2} - \frac{i}{2} - \frac{1}{8}\right)f(i+1) + \left(\frac{3i^2}{2} + \frac{i}{2} - \frac{1}{8}\right)f(i-1)\right]$$

$$+ (i^2 + \tfrac{1}{12})\,[6if(i) + (\tfrac{1}{2} - 3i)\,f(i+1) - (\tfrac{1}{2} + 3i)\,f(i-1)]$$

$$+\left(i^3 + \frac{i}{4}\right)[-3f(i) + \tfrac{3}{2}f(i+1) + \tfrac{3}{2}f(i-1)]\Big\}$$

$$+ \{-m\,[\tfrac{3}{4}\,(m^2 + m - \tfrac{1}{12})\,f(-m+1) - \tfrac{3}{4}\,(m^2 + m - \tfrac{17}{12})\,f(-m)]$$

$$+ (m^2 + \tfrac{1}{12})\,[\tfrac{3}{4}\,(2m+1)\,f(-m+1) - \tfrac{3}{4}\,(2m+1)\,f(-m)]$$

$$-\left(m^3 + \frac{m}{4}\right)[\tfrac{3}{4}f(-m+1) - \tfrac{3}{4}f(-m)]\}$$

$$+ \{n[\tfrac{3}{4}\,(n^2 + n - \tfrac{1}{12})\,f(n-1) - \tfrac{3}{4}\,(n^2 + n - \tfrac{17}{12})\,f(n)]$$

$$+\left(n^2 + \frac{1}{12}\right)\left[-\frac{3}{4}\,(1+2n)\,f(n-1) + \frac{3}{4}\,(1+2n)\,f(n)\right]$$

$$+\left(n^3 + \frac{n}{4}\right)\left[\frac{3}{4}f(n-1) - \frac{3}{4}f(n)\right]\Big\}.$$

$$\bar{\nu}_1 = \sum_{-m+1}^{n-1}\left[if(i) + \frac{1}{24}f(i+1) - \frac{1}{24}f(i-1)\right] + \frac{1}{16}f(-m+1)$$

$$-\left(m + \frac{1}{16}\right)f(-m) - \frac{1}{16}f(n-1) + \left(n + \frac{1}{16}\right)f(n).$$

$$\sum_{-m+1} if(i) = \sum_{-m} if(i) - (-m)\,f(-m) - nf(n) = \nu_1 + mf(-m) - nf(n).$$

$$\frac{1}{24}\sum_{i=-m+1}^{n-1} f(i+1) = \frac{1}{24}\sum_{k=-m+2}^{n} f(k) = \frac{1}{24}\sum_{-m}^{n} f(k) - \frac{1}{24}f(-m+1) - \frac{1}{24}f(-m)$$

$$= \frac{1}{24} - \frac{1}{24}f(-m+1) - \frac{1}{24}f(-m).$$

$$\frac{1}{24}\sum_{i=-m+1}^{n-1} f(i-1) = \frac{1}{24}\sum_{j=-m}^{n-2} f(j) = \frac{1}{24}\sum_{-m}^{n} f(j) - \frac{1}{24}f(n-1) - \frac{1}{24}f(n)$$

$$= \frac{1}{24} - \frac{1}{24}f(n-1) - \frac{1}{24}f(n)\,.$$

$$\bar{\nu}_1 = \nu_1 + mf(-m) - nf(n) + \frac{1}{24} - \frac{1}{24}f(-m+1) - \frac{1}{24}f(-m) - \frac{1}{24}$$

$$+ \frac{1}{24}f(n-1) + \frac{1}{24}f(n) + \frac{1}{16}f(-m+1)$$

$$- \left(m + \frac{1}{16}\right)f(-m) - \frac{1}{16}f(n-1) + \left(n + \frac{1}{16}\right)f(n)\,.$$

$$\bar{\nu}_1 = \nu_1 - \frac{5}{48}f(-m) + \frac{5}{48}f(n) + \frac{1}{48}f(-m+1) - \frac{1}{48}f(n-1)\,.$$

Using this same notation and method for the higher moments we get

$$\bar{\mu}_2 = \nu_2 - \frac{1}{12} - \bar{\nu}_1^2 + \left(\frac{5m}{24} + \frac{7}{80}\right)f(-m) + \left(\frac{5n}{24} + \frac{7}{80}\right)f(n)$$

$$+ \left(\frac{-m}{24} - \frac{1}{240}\right)f(-m+1) + \left(\frac{-n}{24} - \frac{1}{240}\right)f(n-1)\,.$$

$$\bar{\mu}_3 = \nu_3 - 3\bar{\nu}_1\bar{\mu}_2 - \frac{\bar{\nu}_1}{4} - \bar{\nu}_1^3 + f(-m)\left[\frac{-5}{16}m^2 - \frac{21}{80}m - \frac{17}{120}\right]$$

$$+ f(n)\left[\frac{5}{16}n^2 + \frac{21}{80}n + \frac{17}{120}\right] + f(-m+1)\left[\frac{m^2}{16} + \frac{m}{80} + \frac{1}{120}\right]$$

$$+ f(n-1)\left[\frac{-n^2}{16} - \frac{n}{80} - \frac{1}{120}\right].$$

$$\bar{\mu}_4 = \nu_4 - 4\bar{\mu}_3\bar{\nu}_1 - 6\bar{\mu}_2\bar{\nu}_1^2 - \bar{\nu}_1^4 - \frac{\bar{\mu}_2}{2} - \frac{\bar{\nu}_1^2}{2} - \frac{17}{64}$$

$$+ f(-m)\left[\frac{5m^3}{12} + \frac{21m^2}{40} + \frac{17m}{30} + \frac{313}{1680}\right] + f(n)\left[\frac{5n^3}{12} + \frac{21n^2}{40} + \frac{17n}{30} + \frac{313}{1680}\right]$$

$$+ f(-m+1)\left[\frac{-m^3}{12} - \frac{m^2}{40} - \frac{m}{30} - \frac{1}{336}\right] + f(n-1)\left[\frac{-n^3}{12} - \frac{n^2}{40} - \frac{n}{30} - \frac{1}{336}\right].$$

SPECIAL CASES

The above formulae are rather long and in practice the special cases below will frequently be preferred.

(*a*) We may usually take the origin at or very near the middle of the range so that $m = n$, at least approximately.

If $m = n$:

$$\bar{\nu}_1 = \nu_1 - \frac{5}{48}f(-m) + \frac{5}{48}f(n) + \frac{1}{48}f(-m+1) - \frac{1}{48}f(n-1).$$

$$\bar{\mu}_2 = \nu_2 - \frac{1}{12} - \bar{\nu}_1^2 + \left(\frac{5m}{24} + \frac{7}{80}\right)[f(-m) + f(n)]$$
$$+ \left(\frac{-m}{24} - \frac{1}{240}\right)[f(-m+1) + f(n-1)].$$

$$\bar{\mu}_3 = \nu_3 - 3\bar{\nu}_1\bar{\mu}_2 - \frac{\bar{\nu}_1}{4} - \bar{\nu}_1^3 + \left[\frac{5m^2}{16} + \frac{21m}{80} + \frac{17}{120}\right][f(n) - f(-m)]$$
$$+ \left[\frac{m^2}{16} + \frac{m}{80} + \frac{1}{120}\right][f(-m+1) - f(n-1)].$$

$$\bar{\mu}_4 = \nu_4 - 4\bar{\mu}_3\bar{\nu}_1 - 6\bar{\mu}_2\bar{\nu}_1^2 - \bar{\nu}_1^4 - \frac{\bar{\mu}_2}{2} - \frac{\bar{\nu}_1^2}{2} - \frac{17}{64}$$
$$+ \left[\frac{-m^3}{12} - \frac{m^2}{40} - \frac{m}{30} - \frac{1}{336}\right][f(-m+1) + f(n-1)]$$
$$+ \left[\frac{5m^3}{12} + \frac{21m^2}{40} + \frac{17m}{30} + \frac{313}{1680}\right][f(-m) + f(n)].$$

(*b*) Except in the abrupt cases the end frequencies and the difference between those next to the ends will be so small (relative to unity) that they will have a negligible effect on the corrections. If $m = n$ as in (*a*), and if also

$$f(-m) = f(n) = 0 \text{ and } f(-m+1) - f(n-1) = 0:$$

$$\bar{\nu}_1 = \nu_1.$$

$$\bar{\mu}_2 = \nu_2 - \bar{\nu}_1^2 - \frac{1}{12} + f(-m+1)\left[\frac{-m}{12} - \frac{1}{120}\right].$$

$$\bar{\mu}_3 = \nu_3 - 3\bar{\nu}_1\bar{\mu}_2 - \frac{\bar{\nu}_1}{4} - \bar{\nu}_1^3.$$

$$\bar{\mu}_4 = \nu_4 - 4\bar{\mu}_3\bar{\nu}_1 - 6\bar{\mu}_2\bar{\nu}_1^2 - \bar{\nu}_1^4 - \frac{\bar{\mu}_2}{2} - \frac{\bar{\nu}_1^2}{2} - \frac{17}{64}$$
$$+ f(-m+1)\left[\frac{-m^3}{6} - \frac{m^2}{20} - \frac{m}{15} - \frac{1}{168}\right].$$

These formulae have been written in the form which makes the computing simple. The following makes a comparison with Sheppard's corrections easy.

$$\bar{\nu}_1 = \nu_1 .$$

$$\bar{\mu}_2 = \mu_2 - \frac{1}{12} + f(-m+1)\left[\frac{-m}{12} - \frac{1}{120}\right].$$

$$\mu_3 = \mu_3 + \nu_1\left(\frac{m}{4} + \frac{1}{40}\right)f(-m+1).$$

$$\bar{\mu}_4 = \mu_4 - \frac{\mu_2}{2} - \frac{43}{192} + f(-m+1)\left[\frac{-m^3}{6} - \frac{m^2}{20} - \frac{m}{40} - \frac{1}{560} - \frac{m\nu_1^2}{2} - \frac{\nu_1^2}{20}\right].$$

The following special case is also useful in comparing my formulae with Sheppard's.

(*c*) Let $f(-m) = \frac{1}{5} f(-m+1)$ and $f(n) = \frac{1}{5} f(n-1)$. This produces a graduating curve which is exactly tangent to the t-axis at the ends of the range and is everywhere continuous—though it does not have continuous derivatives at certain isolated points. It is, however, a curve which to the eye cannot be distinguished from the type assumed in the Euler-MacLaurin theorem, which lies at the base of Sheppard's formulae. My corrections become:

$$\bar{\nu}_1 = \nu_1 ,$$

$$\bar{\mu}_2 = \mu_2 - \frac{1}{12} + \frac{1}{15}[f(-m) + f(n)],$$

$$\bar{\mu}_3 = \mu_3 - \frac{\nu_1}{5}[f(-m) + f(n)] + \left[\frac{-m}{5} - \frac{1}{10}\right]f(-m) + \left[\frac{n}{5} + \frac{1}{10}\right]f(n),$$

$$\bar{\mu}_4 = \mu_4 - \frac{\mu_2}{2} - \frac{43}{192} + \frac{2}{5}\left[(\nu_1 + m)^2 + \nu_1 + m + \frac{29}{84}\right]f(-m)$$
$$+ \frac{2}{5}\left[(\nu_1 - n)^2 - \nu_1 + n + \frac{29}{84}\right]f(n).$$

Sheppard's are:

$$\bar{\nu}_1 = \nu_1 ,$$

$$\bar{\mu}_2 = \mu_2 - \frac{1}{12},$$

$$\bar{\mu}_3 = \mu_3 ,$$

$$\bar{\mu}_4 = \mu_4 - \frac{\mu_2}{2} + \frac{7}{240}.$$

Let us compare my results with Sheppard's in the very special case in which $f(-m) = f(n) = 1/7$, $f(0) = 5/7$, $m = n = 1$. The odd moments vanish. My corrections for μ_2 and μ_4 are

$$\bar{\mu}_2 = 0\ 2214, \qquad \bar{\mu}_4 = 0.1870.$$

Sheppard's are

$$\bar{\mu}_2 = 0.2024, \qquad \bar{\mu}_4 = 0.1720.$$

The numerical difference between the $\bar{\mu}_2$'s is 0 0190, and the numerical difference between the $\bar{\mu}_4$'s is 0 0150.

This example shows that Sheppard's corrections are not valid to the precision to which they are usually given if they are to be used for the purpose of correcting raw moments. The last term in the fourth moment correction, 7/240, might equally well be, for example, $-43/192$ as in my special case. This will become more evident to the reader if he will draw the curve indicated in this example. To the eye it will appear exactly like the kind specified in the Euler-MacLaurin theorem; for example, much like the normal curve. Now suppose one adopted for the moment the point of view (which I have criticized earlier) of starting with the curve used in this example, breaking it up into three partial areas and then finding the relation between the true and the raw moments. The partial areas found would be exactly those used in this example and this method would give us Sheppard's corrections, but they would not be exactly correct, for in this instance my formulae give exactly the relationship between the true and the raw moments. The difference is due to the fact that in this instance the assumptions permitting the use of the Euler-MacLaurin theorem in abbreviated form are not justified for this curve. But there is no way of telling at the outset, if one has given initially only the partial areas, whether precisely this curve or another which to the eye would appear very much like it is truly the curve which will graduate the same material when subjected to a finer classification.

THE POINT BINOMIAL AND PROBABILITY PAPER

By Frank H. Byron[1]

1. An approximation to the sum of a number of consecutive terms of the point binomial may be found graphically and quite expeditiously by means of so-called "probability paper." This paper is ruled so that the (x, y) graph of the equation of the integral of the normal curve

$$y = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \quad ^{\overline{2}}\, dx \tag{1}$$

is a straight line. Let the successive terms of the point binomial be represented as follows:

$$(p + q)^n = u_0 + u_1 + \cdots + u_t + \cdots + u_n\,, \tag{2}$$

where $u_t = {}_nC_t p^{n-t} q^t$ and $p \geqq q$. Then the (x, y) graph of the equation,

$$y = \sum_{\iota=0}^{t} u_\iota\,, \qquad\qquad t + \tfrac{1}{2} = x\,, \tag{3}$$

i.e., of the sum of first $(t + 1)$ terms of this point binomial, is, in all but extreme cases, a set of points lying on a gently turning curve, so gently that its form may be represented closely by two straight lines, each passing through the median point as will be explained in the next section. As paper of this sort is readily obtainable, and as this method yields as great accuracy as is really useful in many problems, it is suggested that its use ought to be quite general.

2. **Sheppard's Corrections.** The formulae for the moments of the point binomial, mean $= qn$, $\sigma^2 = pqn$, are exact without any corrections such as are used for grouped material. This fact has led us all (apparently) to assume that in fitting the curve to the point binomial one would get a better fit by equating the moments of the curve to the uncorrected moments of the point binomial rather than to the corrected moments. The studies made in connection with the preparation of this paper show that when the purpose is to equate areas to sums of terms the corrected moments should be used. The theoretical basis for this conclusion is as follows:

To simplify the argument let us suppose that one were seeking that curve of Charlier type,

$$F(x) = c_0\phi_0(x) + c_1\phi_1(x) + \cdots c_4\phi_4(x)\,, \tag{4}$$

[1] With the assistance of Burton H. Camp.

(where ϕ_0 is the normal curve and ϕ_1, ϕ_2, $\cdots$ its successive derivatives) whose integral would best fit the graph of (3). Since fitting is required only at the isolated points $x = \frac{1}{2}, 1\frac{1}{2}, 2\frac{1}{2}, \cdots$, it is clear that one might obtain this by the two following steps. First let $f(x)$ be any function whose integral meets exactly the requirement at these isolated points. What values this integral has at other points does not for the moment concern us. There are an infinite number of such $f(x)$ curves. Next let the c's of (4) be so chosen that $F(x)$ will fit $f(x)$ as nearly as possible. The ordinary derivation of the c's supposes that the fit between $f(x)$ and $F(x)$ is to be made by least squares, the residuals being weighted by the factor $1/\sqrt{\phi(x)}$. No matter what $f(x)$ is chosen, the c's can be determined so that the weighted integral of $(f(x) - F(x))^2$ will be a minimum, but the value of this minimum will vary from one $f(x)$ to another. We now desire to select that $f(x)$ which will make this minimum value as small as possible, and it is reasonable to suppose that our best selection will be some $f(x)$ which is as kindred to the nature of $F(x)$ as possible. We shall not therefore choose an $f(x)$ which oscillates wildly between the points where perfect fitting is required, (Fig. 1) nor yet an $f(x)$ which is made up of the top bases of the point binomial

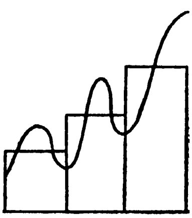


FIG 1

histogram; we shall prefer a modification (Fig. 2) of that histogram by a smoothing process. Such an $f(x)$ will not have the exact moments of the point binomial, but, more nearly, those moments corrected for grouping. Then the determination of the c's will come out in terms of these corrected moments, not in terms of the uncorrected moments. (In fact the uncorrected moments would be the exact moments of an $f(x)$ having an oscillatory character between the important points.)

Of course, when n is large, the difference is too small to be noticed and the use of Sheppard's corrections is not worth while, and since n usually is large when approximations of this sort are needed, the point is not usually important. It was important in the computation of the tables of §4. Moreover, the use of Sheppard's corrections does not invariably yield better results, the gain being masked sometimes by other effects to be considered in §3. An excellent illustration of uniformly better results is in fitting $(\frac{1}{2} + \frac{1}{2})^9$ by a curve of Type 4. The errors in the sums as derived from (4) with and without the corrections, is given on the following page.

t	0	1	2	3	4	5	6	7	8	9
With Corrections	0002	0001	− 0003	− 0001	0000	0001	0003	− 0001	− 0002	0000
Without Corrections	0007	0022	0039	0036	0000	− 0036	− 0039	− 0022	− 0007	− 0001

3. **The Stubby End.** The other effects which mask this improvement are especially noticeable at the stubby end of a point binomial. We have to keep in mind here that the approximating curve (such as (4)), is required to turn a sharp corner, for, due to the least square method of fitting, it is just as important that it be close to zero when t is negative, as it is that it be close to u_0, u_1, $\cdots$ when t is positive. Therefore, in order to turn this corner it has to dip below the x-axis in the neighborhood of $t = -\frac{1}{2}$. This makes the approximating curve too low just to the right of $t = -\frac{1}{2}$, unless the whole curve be arbitrarily widened. This arbitrary widening is customarily performed by not using Sheppard's correction for σ, and the result is a betterment of the fit at these points but a corresponding loss over the rest of the infinite interval. A good example[2] is $(\frac{2}{3} + \frac{1}{3})^{25}$. The fit is worse at the left end when Sheppard's corrections are used but better over the rest of the interval.

The same difficulty arises in another connection. If we compare the closeness of fit to a point binomial made by $F(x)$ as written in (4) and by $F(x)$ as it would be written if c_4 were zero, it often happens (as is well known) that the latter is actually slightly better on the average. How can this be true if the c's are chosen by the method of least squares and the best choice as thus indicated makes c_4 different from zero? The answer is that the c's are chosen so that the fit is best over the infinite interval, not merely over the interval from $t = -\frac{1}{2}$ to $t = n + \frac{1}{2}$, and that furthermore the distant points are weighted more heavily than those near the center. Thus it might happen that a choice, other than the least square choice, and one in which c_4 would be zero, might be better for the restricted interval covered by the point binomial. This does happen especially when due to the abruptness of the stubby end of a very skew binomial, the curve has to dip below the axis in order to get by a sharp corner. A good example is the problem considered by Fry:[3] $(\frac{9}{10} + \frac{1}{10})^{100}$. All the effects mentioned are present here. The fit is on the average a little worse if c_4 is not equal to zero over the point binomial interval, a little better over the infinite interval.

4. For graphical purposes a sufficiently good approximation to the median of $(p + q)^n$, is given by

$$M = nq - (p - q)/6.$$

[2] The true values are given on page 220 of Mathematical Part of Elementary Statistics, by Camp, D. C. Heath and Company, 1931.

[3] T. C. Fry, Probability and its Engineering Uses, p. 258, Van Nostrand, 1928.

The following tables enable us to find the first quartile Q_1, and the ninth decile D_9. The accuracy to which they can be plotted is only about one-tenth that to which they are given here. Therefore accurate interpolation is seldom necessary. The values of S_{t+1} are to be read from the graph at the points $t + \frac{1}{2}$, as indicated in the directions preceding the tables. The graphical method will be found efficient if one uses common sense in the computation. Numbers which are to be plotted should not be computed to a higher degree of accuracy than can be used graphically. In reading the values of S_{t+1} it is well to remember that the true values lie on a curve, and that outside the interval from Q_1 to D_9, they are slightly less than those given by the straight line. Once the graph has

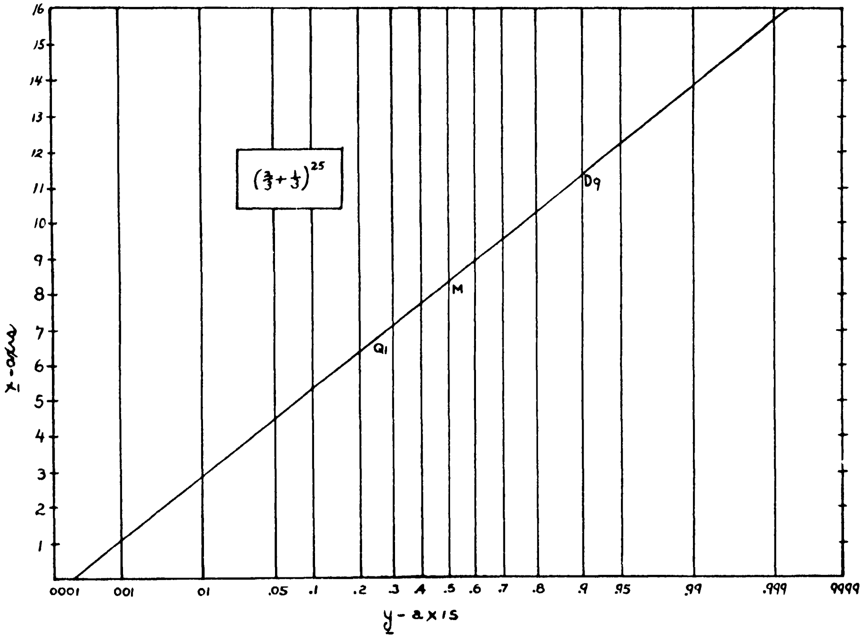


FIG. 3

been made, all the values of S_{t+1} can be read quickly; it is not necessary to make a separate computation for each t. This method is therefore specially advantageous when one wishes to find several sums of this sort for the same point binomial. It should also be noticed that one can tell from the appearance of the graph about how far the true sum would be from the two straight lines and so estimate the error to which his reading is liable.

5. **Illustration.** Find the sum of the first 7 terms of $(\frac{2}{3} + \frac{1}{3})^{25}$.

Here $t = 6$, $M = 8.278$, $Q_1 = 6.726$, $D_9 = 11.369$. The graph shows that $\sum_0^t = 0.224$. The true value is 0.222. So the error is 0.002.

An idea of the accuracy of the method is given by the errors (out of two places) that would be obtained for this point binomial for various values of t, as follows:

					10	12	14	16
Errors	00	01	00	00	00	00	00	00

Directions for Use of the Tables: Let $p = q$, $M = nq - (p - q)/6$, $Q_1 = x_1 + qn$, $D_9 = x_2 + qn$. On the graph draw the lines MQ_1 and MD_9. Read S_{t+1} at $t + \frac{1}{2}$.

Values of x_1

	2000	1000	750	500	400	300	200	100	75	50	25
99	−693	−701	−704	−710	−714	−720	−728	−747	−756	−771	−804
98	−688	−693	−696	−700	−703	−707	−714	−728	−735	−746	−770
97	−685	−690	−692	−696	−698	−701	−707	−718	−.724	−734	−784
96	−684	−687	−689	−693	−695	−697	−702	−712	−718	−726	−744
95	−683	−686	−688	−691	−692	−695	−699	−708	−713	−721	−737
94	−682	−685	−686	−689	−691	−693	−697	−705	−709	−717	−732
93	−681	−684	−685	−688	−689	−691	−695	−703	−707	−713	−727
92	−681	−683	−685	−687	−688	−690	−693	−701	−704	−710	−723
91	−680	−683	−684	−686	−687	−689	−692	−699	−702	−708	−720
90	−680	−682	−683	−685	−686	−688	−690	−697	−700	−704	−717
88	−679	−681	−682	−684	−685	−686	−689	−695	−697	−702	−713
85	−679	−680	−681	−682	−683	−685	−687	−691	−694	−698	−707
80	−677	−679	−679	−681	−681	−682	−684	−688	−690	−693	−700
75	−677	−678	−678	−679	−680	−681	−682	−685	−686	−689	−694
70	−676	−677	−677	−678	−679	−679	−680	−682	−683	−685	−690
65	−676	−676	−677	−677	−677	−678	−678	−680	−681	−682	−686
60	−675	−676	−676	−676	−676	−677	−677	−678	−679	−680	−682
50	−675	−675	−675	−675	−675	−675	−675	−675	−675	−675	−675

ERRATA

The Annals of Mathematical Statistics

Volume VI, No. 1, March, 1935

On page 25, in [illegible]ions for Use of the Tables, $p = q$ should read $p \gtreqless q$, $Q_1 = x_1 + qn$ should read $Q_1 = x_{1\sigma} + qn$; $D_9 = x_2 + qn$ should read $D_9 = x_{2\sigma} + qn$. In the tables of values of x under $p = .97$, $n = 25$, instead of $-.784$ the number should be $-.754$.

Values of x_2

p \ n	2000	1000	750	500	400	300	200	100	75	50	25
99	1 307	1 318	1 325	1 336	1 344	1 356	1 378	1 439	1 481	See Auxiliary	
98	299	307	311	318	323	330	343	376	396	Tables	
97	295	301	304	310	314	319	329	353	367		
96	293	298	301	306	309	313	321	341	352		
.95	292	296	299	303	305	309	316	332	342		
94	291	295	297	300	303	306	312	327	335		
93	290	293	295	298	301	304	309	322	329	1 342	1 374
92	289	292	294	297	299	302	307	318	325	336	365
.91	289	292	293	296	298	300	305	315	321	331	357
90	288	291	292	295	296	299	303	313	318	325	351
88	287	290	291	293	295	297	300	309	313	321	341
85	286	288	289	291	292	294	297	304	308	314	330
.80	285	287	288	289	290	291	293	298	301	306	317
75	284	285	286	287	288	289	291	294	297	300	308
.70	284	285	285	286	286	287	288	291	293	295	301
65	283	284	284	285	285	286	286	288	290	292	296
.60	283	283	283	284	284	284	285	286	287	288	291
.50	282	282	282	282	282	282	282	282	282	282	282

Auxiliary Table

p \ n	60	50	40	35	30	25	20
99	1 525	1 575	1 663	1 740	1.871	2 149	3 209
98	416	435	455	488	520	1 568	1 652
97	381	394	413	433	445	472	514
96	362	372	387	397	410	428	457
.95	350	359	370	378	389	405	425
94	336	340	359	366	375	387	405

INEQUALITIES AMONG AVERAGES

By Nilan Norris

Numerous inequalities among averages of various types are condensed in the monotonic character of the function

$$\phi(t) = \left(\frac{x_1^t + x_2^t + \cdots + x_n^t}{n}\right)^{\frac{1}{t}}$$

of the positive numbers $x_1, x_2, \cdots, x_n$, not all equal each to each. For $t = -1$ this function is the harmonic mean; for $t = 0$ it is the geometric mean; for $t = 1$ the arithmetic mean; and for $t = 2$ the root mean square. The relations among these four means which customarily are proved by special and disconnected methods appear easily as applications of the theorem that $\phi(t)$ is an increasing function of t. That is, for any values of t_1 and t_2 such that $-\infty < t_1 < t_2 < +\infty$, it will be true that $\phi(t_1) < \phi(t_2)$. Several proofs of this theorem have been published, many of them very complex. An extremely simple proof is herewith presented.[1]

That $\phi(t)$, $\phi'(t)$ and $\phi''(t)$ all exist and are continuous for all real values of t may be shown by expanding each of the quantities x_i^t in a series of powers of t and considering the remainders after each of the first three terms. The ordinary rule for evaluating forms reducing to $0/0$, which requires the function under consideration to be continuous and to have at least a continuous first derivative for $t = 0$, may then be applied to $[\log \phi(t)]/t$ to show that $\phi(0)$ is the geometric mean. It is clear that $\phi(-\infty)$ and $\phi(+\infty)$ are respectively the least and the greatest of the x_i. This fact and the monotonic property of $\phi(t)$ make it evident that for each real value of t, the function may be regarded as an average in the usual sense that it lies within the range of the observations.

For a simple demonstration of the increasing character of $\phi(t)$, consider the auxiliary function

$$F(t) = t^2 \frac{\phi'(t)}{\phi(t)} = t^2 \frac{d}{dt}\left\{\frac{1}{t} \log \frac{\Sigma x^t}{n}\right\} = t \frac{\Sigma x^t \log x}{\Sigma x^t} - \log \frac{\Sigma x^t}{n}.$$

It is clear that $\phi'(t)$ has the same sign as $F(t)$. The theorem will be proved by showing that the sign of $F(t)$ is positive for all values of t except zero, when $\phi'(t)$ vanishes.

[1] Professor Harold Hotelling rendered invaluable assistance in condensing for publication the material herein presented from a more extended study of generalized mean value functions

Differentiating the last expression with respect to t, one obtains upon simplification

$$F'(t) = \frac{t}{(\Sigma\, x^t)^2}\,[(\Sigma x^t)\,(\Sigma\, x^t \log^2 x) - (\Sigma x^t \log x)^2]\,.$$

By Cauchy's inequality (known as Schwarz' inequality when applied to integrals instead of sums), the expression in square brackets is positive. Hence $F'(t)$ has the same sign as t. Consequently $F(t)$, since it diminishes for negative values of t and increases for positive values, has a minimum for $t = 0$. But by direct substitution, $F(0) = 0$. It follows that $F(t)$ and $\phi'(t)$ are positive for all values of t other than zero. Therefore $\phi(t)$ is an increasing function.

By direct general methods it is possible to show that

$$\phi'(0) = (\Pi x)^{\frac{1}{n}}\,\frac{1}{2n^2}\,[n\Sigma(\log x)^2 - (\Sigma \log x)^2].$$

This expression obviously vanishes only when $n\Sigma(\log x)^2 = (\Sigma \log x)^2$, a condition which is satisfied only in the trivial case when $x_1 = x_2 = \cdots = x_n$.

A proof exactly parallel to that given above may be applied to integrals or, more generally, to Stieltjes integrals. The monotonic increasing character of $\left[\int_{x=0}^{\infty} x^t d\psi(x)\right]^{\frac{1}{t}}$ appears in this way if one assumes that $\psi(x)$ is a non-decreasing function integrable in the Riemann-Stieltjes sense, such that $\psi(\infty) - \psi(0) = 1$, and such that $\int_{x=0}^{\infty} x^t d\psi(x)$ exists for every real value of t. In terms of statistical theory, this consideration extends the theorem from samples to populations of a very general character.

Proof of the increasing character of $\phi(t)$ has also been derived from Hölder's inequality, the demonstration being expressed in terms of Stieltjes integrals.[2] The simplest general proof of the monotonic attribute of $\phi(t)$ heretofore published appears to be that of Paul Lévy.[3] As early as 1840 Bienaymé[4] presented a generalized form of $\phi(t)$, namely,

$$\left(\frac{c_1 a_1^m + c_2 a_2^m + \cdots + c_n a_n^m}{c_1 + c_2 + \cdots + c_n}\right)^{\frac{1}{m}},$$

and announced, without proof, its increasing character. In 1858 a proof of the monotonic quality of $\phi(t)$ for special cases was published by Schlömilch.[5] Of

[2] J. Shohat, "Stieltjes Integrals in Mathematical Statistics," *Annals of Mathematical Statistics* (American Statistical Association, Ann Arbor, 1930), Vol. 1, No. 1, p. 84.

[3] *Calcul des Probabilités* (Gauthier-Villars et Cie., Paris, 1925), pp. 157 *f*.

[4] Jules Bienaymé, *Société Philomatique de Paris*, Extraits des Procès-Verbaux des Seances Pedant L'Anée 1840 (Imprimerie D'A. René et Cie., Paris, 1841), Seance du 13 juin 1840, p. 68.

[5] O. Schlömilch, "Ueber Mittelgrössen verschiedener Ordnungen," *Zeitschrift für Mathematik und Physik* (B. G. Teubner, Leipzig, 1858), Vol. 3, pp. 303 *f*.

the more recent general proofs of the increasing character of $\phi(t)$ which have appeared, those of Jensen,[6] Pólya,[7] Jessen,[8] and Carathéodory[9] may be mentioned. A recent application of $\phi(t)$ to index number theory is that of Professor John B. Canning.[10]

VASSAR COLLEGE.

[6] J. L. W. V. Jensen, "Sur Les Fonctions Convexes Et Les Inegalités Entre Les Valeurs Moyennes," *Acta Mathematica* (Beijers Bokforlagsaktielbolag, Stockholm, 1905), Vol. 30, pp. 183–185.

[7] G. Pólya and G. Szego, *Aufgaben und Lehrsatze Aus Der Analysis* (Julius Springer, Berlin, 1925), Vol. I, pp. 54 *f*. and 210.

[8] Børge Jessen, "Bemaerkninger om koveskse Funktioner og Uligheder imellem Middelvaerdier," *Matematisk Tidsskrift* (Charles Johansens Bogtrykkeri, Copenhagen, 1931), No. 2, 1931, pp. 26–28.

[9] Attributed to Professor Constantin Carathéodory in an unpublished manuscript of Professor Harold Hotelling

[10] "A Theorem Concerning a Certain Family of Averages of a Certain Type of Frequency Distribution," a paper presented before a joint meeting of the American Statistical Association and the Econometric Society at Berkeley, California, June 22, 1934.

MATHEMATICAL EXPECTATION OF PRODUCT MOMENTS OF SAMPLES DRAWN FROM A SET OF INFINITE POPULATIONS

By Hyman M. Feldman[1]

Introduction

In the second part of his investigations, "On the Mathematical Expectation of Moments of Frequency Distributions,"[2] Tchouproff presented a method which may be interpreted as sampling from a set of infinite univariate populations. In the present paper this method is extended to the study of moments of product moments of samples drawn from a set of infinite bivariate populations. It is also shown how this method may be extended to populations of higher order by deriving some of the simpler formulae for populations of three and four variables.

Tchouproff's method has been criticised[3] because of the complicated algebra. On close examination it is found, however, that it is not the algebra which is complicated but rather the symbolism. Tchouproff introduced a great variety of symbols both in his derivations and in his results. As a consequence his work seems very intricate. If, however, the number of symbols is reduced, and the symbols themselves are simplified, which can be easily accomplished, the underlying idea of Tchouproff's method is found to be very simple.

Quite a complete study of product moments of any bivariate population has been made by Joseph Pepper in his "Studies in the Theory of Sampling."[4] His method is essentially an extension of Church's[5] method, in his studies of univariate populations, to bivariate populations. He does not, however, derive any generalized formulae. In the present study generalized formulae for both the first moment and the variance of product moments of any order are obtained.

It may be noted here, that all of Pepper's formulae for any infinite population can be obtained from those of the present study as special cases, by assuming that all the populations in the set are identical.

[1] A dissertation presented to the Board of Graduate Studies of Washington University in partial fulfilment of the requirements for the degree of Doctor of Philosophy, June 1933.

[2] *Biometrika*, Vol. XXI, Dec. 1929, pp. 231–258.

[3] Church, A. E. R. "On the Means and Squared Standard Deviations of Small Samples from any Population," *Biometrika*, Vol. XVIII, Nov., 1926, pp. 321–394.

[4] *Biometrika*, Vol. XXI, Dec. 1929, pp. 231–258.

[5] Church, A. E. R., "On the Means and Squared Standard Deviations of Small Samples rom any Population," *Biometrika*, Vol. XVIII, Nov., 1926, pp. 321–394.

Chapter I. Notations and Definitions

Let $(X_1, Y_1), (X_2, Y_2), \cdots (X_n, Y_n)$ be n bivariate populations each following any law of distribution whatever. The product moment of order a in X and b in Y of the k^{th} population will be denoted by $P^k_{a\,b}$. It is defined as

$$P^k_{a\,b} = E(X_k - a_k)^a \, (Y_k - b_k)^b \tag{1.11}$$

where
$$a_k = E(X_k), \qquad b_k = E(Y_k), \tag{1.12}$$

and where the symbol E signifies the expected value or the mathematical expectation of a quantity.

Regarding each of the n populations of the set as infinite,[6] samples of n are drawn, each member of a sample from one of the n populations.[7] The individual which is drawn from the k^{th} population will be denoted by (x_k, y_k); and the product moment of order a in x and b in y, of such a sample will be denoted by p_{ab}. This product moment may then be defined as

$$p_{ab} = n^{-1} \, S \, (x_k - x)^a \, (y_k - y)^b \tag{1.13}$$

where
$$x = n^{-1} \, Sx_k, \qquad y = n^{-1} \, Sy_k \, . \tag{1.14}$$

The symbols a and b will now be defined by the equations

$$a = n^{-1} \, Sa_k, \qquad b = n^{-1} \, Sb_k \, . \tag{1.15}$$

Obviously
$$E(x) = E(n^{-1} \, Sx_k) = n^{-1} \, SE(X_k) = n^{-1} \, Sa_k = a. \tag{1.16}$$

Similarly $E(y) = b$. That is, the mathematical expectation of the mean, of such a sample as was described above, is equal to the average of the means of all the populations.[8]

In order to make the equations as compact as possible the following additional symbols will be employed:

$$\begin{aligned} x_k - a_k &= u_k, \qquad & x - a &= u, \qquad & \text{and } u_k - u &= U_k \\ y_k - b_k &= v_k, \qquad & y - b &= v, \qquad & \text{and } v_k - v &= V_k \end{aligned} \tag{1.17}$$

also $a_k - a = A_k, b_k - b = B_k$.

From the above definitions it easily follows that

$$E(u_k) = E(v_k) = E(U_k) = E(V_k) = E(u) = E(v) = 0 \, . \tag{1.18}$$

[6] The term infinite is used here in the probability sense. It is defined very clearly by Church in his "Means and Squared Standard Deviations of Small Samples," *Biometrika*, Vol. XVIII, Nov., 1926, p. 322.

[7] It may be easily shown that this is equivalent to drawing a sample of n from a set of any finite number of populations. The number drawn from each population, however, must be specified. See *Biometrika*, Vol. XIII, 1920–21, p. 295, footnote.

[8] This, of course, is a result of the Lexis Theory, for Poisson and Lexis Series.

The notation is now completed with the definition of the symbol Q_{ij} by the equation:

$$Q_{ij} = S(a_k - a)^i (b_k - b)^j = SA_k^i B_k^j . \tag{1.19}$$

CHAPTER II. The Mathematical Expectation of p_{ab}

The mathematical expectation of p_{ab} will be denoted by $\bar{p}_{ab}$. In the terminology of moments this would be called the mean or first moment of the distribution of p_{ab}.

1. The Mathematical Expectation of p_{11}. According to the above notation the expected value of p_{11} is $\bar{p}_{11}$. By definition

$$\bar{p}_{11} = E(p_{11}) = En^{-1}S(x_i - x)(y_i - y), \tag{2.11}$$

and obviously $En^{-1}S(x_i - x)(y_i - y) = n^{-1}SE(x_i - x)(y_i - y)$.
Writing

$$x_i - x = [(x_i - a_i) - (x - a)] + [a_i - a] = U_i + A_i$$

$$y_i - y = [(y_i - b_i) - (y - b)] + [b_i - b] = V_i + B_i,$$

equation (2.11) may be written as

$$\bar{p}_{11} = n^{-1}SE(U_i + A_i)(V_i + B_i)$$

$$= n^{-1}SE(U_iV_i) + n^{-1}SA_iE(V_i) + n^{-1}SB_iE(U_i) + n^{-1}SE(A_iB_i).$$

Since for any given population A_i and B_i are constants, it follows that $E(A_iB_i) = A_iB_i$. Hence

$$n^{-1}SE(A_iB_i) = n^{-1}SA_iB_i = n^{-1}Q_{11} .$$

Making use of (1.18), it is seen that the terms $n^{-1}SA_iE(V_i)$ and $n^{-1}SB_iE(U_i)$ are zero. The only term left to evaluate is therefore $n^{-1}SE(U_iV_i)$. Since U_i and V_i are symmetric functions of the corresponding small letters, their product is symmetric in u_iv_i. There is therefore no loss in generality if attention is concentrated on a single subscript, say 1.

We may therefore write

$$n^{-1}SE(U_iV_i) = n^{-1}E(U_1V_1) + n^{-1}\underset{*2}{S}E(U_iV_i).$$

Remembering that $U_i = u_i - u = u_i - n^{-1}Su_i$, we may write,

$$U_i = u_i - u = u_i - n^{-1}(u_1 + u_2 + \cdots + u_n)$$

$$= n^{-1}[n_1u_i - (u_1 + u_2 + \cdots + u_{i-1} + u_{i+1} + \cdots + u_n)]$$

* The 2 at the bottom of the S simply indicates that the summation begins with $i = 2$.

where $n_1 = n - 1$. In general, n_i will denote the number $n - i$. Similarly

$$V_i = n^{-1}[n_1 v_1 - (v_1 + v_2 + \cdots + v_{i-1} + v_{i+1} + \cdots + v_n)] .$$

Thus

$$n^{-1}SE(U_iV_i) = n^{-3}E(n_1u_1 - u_2 - \cdots - u_n)(n_1v_1 - v_2 - \cdots - v_n)$$

$$+ n^{-3}\underset{2}{S}E(n_1u_i - u_1 - \cdots - u_{i-1} - u_{i+1} - \cdots u_n)$$

$$(n_1v_i - v_1 - \cdots - v_{i-1} - v_{i+1} - \cdots v_n) .$$

When the right hand side of the last equation is expanded the only terms which appear are of the form $E(u_iv_i)$ and $E(u_iv_j)$. The last one must vanish for u_i and v_j are independent and hence $E(u_iv_j) = E(u_i)E(v_j) = 0$. From the last equation above it is easily seen that the coefficient of $E(u_1v_1)$ is

$$n^{-3}(n_1^2 + n_1) = n^{-3}\, n_1(n_1 + 1) = n^{-2}\, n_1;$$

and because of the symmetry this is obviously the coefficient of any term of that form. Hence

$$n^{-1}SE(U_iV_i) = n^{-2}n_1SE(u_iv_i) .$$

Since $u_i = x_i - a_i$, $v_i = y_i - b_i$, then

$$E(u_iv_i) = E(x_i - a_i)(y_i - b_i) = E(X_i - a_i)(Y_i - b_i) = P^i_{11}$$

and in general,

$$E(u_k^iv_k^j) = P^k_{ij} . \tag{2.12}$$

We thus get the formula

$$\bar{p}_{11} = n^{-2}n_1SP^i_{11} + n^{-1}Q_{11} . \tag{1}$$

Now suppose all the n populations are identical. Then all the A's and also all the B's vanish and therefore, $Q_{11} = 0$. The formula (1) thus becomes

$$\bar{p}_{11} = \frac{n-1}{n} P_{11} . \tag{1'}$$

This is exactly Pepper's formula for $\bar{p}_{11}$ for an infinite population.[9]

2. The Mathematical Expectation of p_{21}. By definition

$$\bar{p}_{21} = En^{-1}S(x_i - x)^2(y_i - y) . \tag{2.21}$$

[9] *Biometrika*, Vol. XXI, p. 233, Eq. A, N = ∞. As was already stated in the introduction, all the formulae of the present study reduce to Pepper's when the above assumption is made.

Proceeding as above it is seen that

$$En^{-1}S(x_\iota - x)^2(y_\iota - y) = n^{-1}SE(x_\iota - x)^2(y_\iota - y)$$
$$= n^{-1}SE(U_\iota + A_\iota)^2(V_\iota + B_\iota) = n^{-1}SE(U_\iota^2 V_\iota) + 2n^{-1}SE(U_\iota V_\iota A_\iota)$$
$$+ n^{-1}SE(U_\iota^2 B_\iota) + n^{-1}SE(V_\iota A_\iota^2) + 2n^{-1}SE(A_\iota B_\iota U_\iota) + n^{-1}SE(A_\iota^2 B_\iota) \cdots \cdot (2.22)$$

It is quite evident that the two terms before the last vanish. To evaluate the remaining terms, we employ the reasoning of section 1 of this chapter and write:

$$SE(U_\iota^2 V_\iota) = E(U_1^2 V_1) + \underset{2}{S}E(U_\iota^2 V_\iota)$$

$$= n^{-3}E(n_1^* u_1 - u_2 - \cdots)(n_1 v_1 - v_2 - \cdots) + n^{-3}\underset{2}{S}E(n_1 u_\iota - u_1 - \cdots)$$
$$(n_1 v_\iota - v_1 - \cdots).$$

Since terms of the form $E(u_\iota^2 v_j)$ vanish, only the coefficient of the term $E(u_\iota^2 v_\iota)$ must be found. Again considering the subscript 1, the coefficient of $E(u_1^2 v_1)$ is easily found from the last equation to be

$$n^{-3}(n_1^3 - n_1) = n^{-3}n_1(n_1 + 1)(n_1 - 1) = n^{-2}n_1 n_2 .$$

Thus

$$n^{-1}SE(U_\iota^2\ V_\iota) = n^{-2}n_1 n_2 SE(u_\iota^2\ v_\iota) = n^{-2}n_1 n_2 SP^{\iota}_{2\,1} . \qquad (2.23)$$

For the second term of (2.22) we have

$$SE(U_\iota V_\iota A_\iota) = E(U_1 V_1 A_1) + \underset{2}{S}E(U_\iota V_\iota A_\iota)$$

$$= n^{-2}E(n_1 u_1 - u_2 - \cdots)(n_1 v_1 - v_2 - \cdots)A_1 + n^{-2}\underset{2}{S}E(n_1 u_\iota - u_1 - \cdots)$$
$$(n_1 v_\iota - v_1 - \cdots)A_\iota.$$

The coefficient of $E(u_1 v_1)$ in the first term of the right hand side of the last equation is $n^{-2}n_1^2 A_1$. In the second term it is $n^{-2}\underset{2}{S}A_\iota = -n^{-2}A_1$, since $SA_\iota = 0$. It therefore follows that

$$2n^{-1}SE(U_\iota V_\iota A_\iota) = 2n^{-2}n_2 SP^{\iota}_{1\,1} A_\iota . \qquad (2.24)$$

Quite similarly

$$n^{-1}SE(U_\iota^2 B_\iota) = n^{-2}n_2 SP^{\iota}_{2\,0} B_\iota, \qquad (2.25)$$

and it is obvious that

$$n^{-1}SE(A_\iota^2 B_\iota) = n^{-1}Q_{21}. \qquad (2.26)$$

* Note that the u which has the coefficient n_1 does not occur among the u's which have the negative sign.

We thus get the formula

$$\bar{p}_{21} = n^{-3} n_1 n_2 S P_{21}^{i} + n^{-2} n_2 S(2 P_{11}^{i} A_i + P_{20}^{i} B_i) + n^{-1} Q_{21}. \qquad (2)$$

3. The Mathematical Expectation of p_{31} and p_{22}.

$$\bar{p}_{31} = E n^{-1} S(x_i - x)^3 (y_i - y) = n^{-1} S E(x_i - x)^3 (y_i - y)$$

$$= n^{-1} S E(U_i + A_i)^3 (V_i + B_i) = n^{-1} S\{E(U_i^3 V_i + U_i^3 B_i + 3U_i^2 V_i A_i$$

$$+ 3U_i^2 A_i B_i + 3U_i V_i A_i^2 + 3U_i A_i^2 B_i + V_i A_i^3 + A_i^3 B_i)\}. \qquad (2.31)$$

The two terms before the last are zero. The last term is

$$n^{-1} S E(A_i^3 B_i) = n^{-1} Q_{31}. \qquad (2.32)$$

By (2.23) and (2.24) and some slight manipulation

$$3n^{-1} S E(U_i^2 A_i B_i + U_i V_i A_i^2)$$

$$= 3n^{-3} n_2 S(P_{20}^{i} A_i B_i + P_{11}^{i} A_i^2) + 3n^{-3}(Q_{11} S P_{20}^{i} + Q_{20}^{i} S P_{11}^{i}), \qquad (2.33)$$

and by (2.22)

$$n^{-1} S E(U_i^3 B_i + 3U_i^2 V_i A_i) = n^{-4}(n_1^3 + 1) S(P_{30}^{i} B_i + 3P_{21}^{i} A_i). \qquad (2.34)$$

The only new term which is to be evaluated is $SE(U_i^3 V_i)$. This may be written as follows:

$$SE(U_i^3 V_i) = n^{-4} S E(n_1 u_i - u_1 - \cdots)^3 (n_1 v_i - v_1 - \cdots).$$

When the right hand side is expanded it is found that the only non-vanishing terms are of the form $E(U_i^3 V_i)$ and $E(u_i^2 u_j v_j)$. Only two subscripts, therefore, have to be considered. Without any loss in generality these may be taken as 1 and 2, and the right hand side of the last equation may then be written as follows:

$$SE(n_1 u_i - u_1 - \cdots)^3 (n_1 v_i - v_1 - \cdots) = E(n_1 u_1 - u_2 - \cdots)^3 (n_1 v_1 - v_2 - \cdots)$$

$$+ E(n_1 u_2 - u_1 - \cdots)^3 (n_1 v_1 - v_2 - \cdots) + \underset{3}{S} E(n_1 u_i - u_1 - u_2 - \cdots)^3$$

$$(n_1 v_i - v_1 - v_2 - \cdots).$$

From this last expansion it is easily seen that the coefficient of $E(u_i^3 v_i)$ is $(n_1^4 + n_1)$ and that of $E(u_i^2 u_j v_j)$, $(6n_1^2 + 3n_2) = 3(2n_1^2 + n_2)$. We thus finally obtain

$$SE(U_i^3 V_i) = n^{-4}\{(n_1^4 + n_1) SE(u_i^3 v_i) + 3(2n_1^2 + n_2) SE(u_i^2 u_j v_j)\}.$$

But by (2.12) $E(u_i^3 v_i) = P_{31}^{i}$, and since u_i and u_j and u_i and v_j are independent $E(u_i^2 u_j v_j) = E(u_i^2) E(u_j v_j) = P_{20}^{i} P_{11}^{j}$. Whence

$$E(U_i^3 V_i) = n^{-4}\{(n_1^4 + n_1) S P_{31}^{i} + 3(2n_1^2 + n_2) S P_{20}^{i} P_{11}^{j}\}. \qquad (2.35)$$

From (2.31) and the succeeding equations we finally get

$$\bar{p}_{31} = n^{-5}\{(n_1^4 + n_1)SP_{3\,1}^{i} + 3(2n_1^2 + n_2)SP^{*}{}_{2\,0}^{i}P_{1\,1}^{j}\}$$
$$+ n^{-4}\{(n_1^3 + 1)S(P_{3\,0}^{i}B_i + 3P_{2\,1}^{i}A_i)\} + 3n^{-3}\{(n_1^2 - 1)S(P_{2\,0}^{i}A_iB_i + P_{1\,1}^{i}A_i^2)$$
$$+ Q_{11}SP_{2\,0}^{i} + Q_{20}SP_{1\,1}^{i}\} + n^{-1}Q_{31}. \tag{3}$$

The derivation of $\bar{p}_{22}$ is so similar to that of $\bar{p}_{31}$, that it would be mere repetition to go through the details again. We shall therefore merely write down the formula for $\bar{p}_{22}$ which is

$$\bar{p}_{22} = n^{-5}\{(n_1^4 + n_1)SP_{2\,2}^{i} + (2n_1^2 + n_2)S(P_{2\,0}^{i}P_{0\,2}^{j} + 4P_{1\,1}^{i}P_{1\,1}^{j})\}$$
$$+ 2n^{-4}\{(n_1^3 + 1)S(P_{2\,1}^{i}B_i + P_{1\,2}^{i}A_i)\} + n^{-3}\{(n_1^2 - 1)S(P_{2\,0}^{i}B_i^2 + 4P_{1\,1}^{i}A_iB_i$$
$$+ P_{0\,2}^{i}A_i^2) + Q_{20}SP_{0\,2}^{i} + Q_{02}SP_{2\,0}^{i} + 4Q_{11}SP_{1\,1}^{i}\} + n^{-1}Q_{22}. \tag{4}$$

4. **The Mathematical Expectation of the General Product Moment p_{ab}.** So far, formulae for the mathematical expectation of p_{ab}, for particular values of a and b, have been derived. The method used in deriving these is, however, perfectly general, and now, that it has been sufficiently illustrated, it can be easily generalized.

By definition we have

$$\bar{p}_{ab} = E[n^{-1}S(x_i - x)^a(y_i - y)^b].$$

Making use of the notation of Chapter I this may be written as

$$n\bar{p}_{ab} = ES(U_i + A_i)^a(V_i + B_i)^b = \mathop{S}_{q,r=0}^{a,b} C_q^a C_r^b \mathop{S}_1^n E(U_i^{a-q}V_i^{b-r}A_i^qB_i^r) \tag{2.41}$$

where

$$C_q^a = \frac{a!}{q!(a-q)!}, \qquad C_r^b = \frac{b!}{r!(b-r)!}.$$

Expressing the U's and V's in terms of the u's and v's and setting $a - q = l$, $b - r = m$; we may write for a particular pair of values q and r:

$$n^{l+m}SE(U_i^lV_i^mA_i^qB_i^r) = SE(n_1u_i - u_1 - \cdots)^l(n_1v_i - v_1 - \cdots)^mA_i^qB_i^r. \tag{2.42}$$

Consider, now, the general term in the expansion of the right hand side of (2.42). It is of the form:

$$\frac{l!m!}{\Pi\alpha_h!\Pi\beta_h!}(-1)^{l+m}(-n_1)^{\alpha_h+\beta_h}E(n_1u_{j_1}^{\alpha_1}\cdots u_{j_k}^{\alpha_k}v_{j_1}^{\beta_1}\cdots v_{j_k}^{\beta_k}A_i^qB_i^r), \tag{2.43}$$

where $\Pi\alpha_h! = \alpha_1!\alpha_2!\cdots\alpha_k!$

* In this case, and also in the formulae that follow, whenever two or more indices appear in a summation, it will be understood that no two of them can have the same value simultaneously.

For particular sets of values $j_1, j_2, \cdots j_k$, $\alpha_1, \alpha_2, \cdots \alpha_k$, and $\beta_1, \beta_2, \cdots \beta_k$, this term will appear in every member of the summation of the right hand side of (2.42), and its coefficient will differ only in the exponent of $(-n_1)$ and in the subscript i of A^qB^r. Because of the symmetry there is no loss in generality if we take for $j_1, j_2, \cdots j_k$, the first k integers. We now break up the summation of the right hand side of (2.42) as follows:

$$\begin{aligned}
&\overset{n}{\underset{1}{S}} E(n_1u_i - u_1 - \cdots)^l(n_1v_i - v_1 - \cdots)^m A_i^q B_i^r \\
&\quad = E(n_1u_1 - u_2 - \cdots)^l(n_1v_1 - v_2 - \cdots)^m A_1^q B_1^r \\
&\quad + E(n_1u_2 - u_1 - \cdots)^l(n_1v_2 - v_1 - \cdots)^m A_2^q B_2^r + \cdots + E(n_1u_k - u_1 - \cdots)^l \\
&\quad (n_1v_k - v_1 - \cdots)^m A_k^q B_k^r + \overset{n}{\underset{i=k+1}{S}} E(n_1u_i - u_1 - \cdots)^l \\
&\quad (n_1v_i - v_1 - \cdots)^m A_i^q B_i^r .
\end{aligned} \tag{2.44}$$

From (2.44) we easily get for the total coefficient (excluding the numerical factor) the expression

$$\overset{k}{\underset{h=1}{S}} (-n_1)^{\alpha_h+\beta_h} A_h^q B_h^r + \overset{n}{\underset{h=k+1}{S}} A_h^q B_h^r .$$

Writing

$$\underset{k+1}{S} A_h^q B_h^r = \underset{1}{S} A_h^q B_h^r - \overset{k}{\underset{1}{S}} A_h^q B_h^r = Q_{qr} - \overset{k}{\underset{1}{S}} A_h^q B_h^r ,$$

the general term, (2.43), together with the total coefficient, may then be written as

$$(-1)^{l+m} \frac{l\,!\; m\,!}{\Pi\alpha_h\,!\;\Pi\beta_h\,!} \left\{ \overset{k}{\underset{h=1}{S}} \left[(-n_1)^{\alpha_h+\beta_h} - 1\right] A_h^q B_h^r + Q_{qr} \right\} \overset{k}{\underset{h=1}{E}} u_h^{\alpha_h} v_h^{\beta_h} .$$

Since u_i and u_j, v_i and v_j, and u_i and v_j are independent while u_i and v_i are not, we have:

I. $E\Pi u_h v_h = \Pi E u_h v_h = \Pi P^h_{\alpha_h\beta_h}$

II. Any term in which $\alpha_h + \beta_h = 1$ must vanish.

From II it follows that the maximum number of subscripts which can appear in any term in the expansion of (2.42), i.e. the upper limit of k, which will be denoted by t, cannot exceed $(l + m)/2$. In fact when $l + m$ is even, $t = (l + m)/2$, while when $l + m$ is odd, t is the largest integer less than $(l + m)/2$.

Making use of (2.41), the equations following it, and the reasoning of the last paragraph, we finally get the formula:

$$\begin{aligned}
n(-n)^{a+b}\bar{p}_{ab} &= (a\,!)\,(b\,!) \overset{n}{\underset{j_h=1}{S}} \overset{a,b}{\underset{q,r=0}{S}} \frac{(-n)^{q+r}}{q\,!\; r\,!} \overset{a-q,\,b-r}{\underset{\alpha_h=0,\,\beta_h=0}{S}} \overset{t}{\underset{k=1}{S}} \\
&\left\{ \overset{k}{\underset{h=1}{S}} \left[(-n_1)^{\alpha_h+\beta_h} - 1\right] A_{j_h}^q B_{j_h}^r + Q_{qr} \right\} \Pi \frac{P^{j_h}_{\alpha_h\beta_h}}{\alpha_h\,!\;\beta_h\,!} .
\end{aligned} \tag{5}$$

The following restrictions on the α's and β's must be observed

$$\text{(a)}\ \alpha_1 + \alpha_2 + \cdots + \alpha_k = a - q$$

$$\text{(b)}\ \beta_1 + \beta_2 + \cdots + \beta_k = b - r$$

$$\text{(c)}\ \alpha_h + \beta_h \neq 1\,.$$

In case the n populations are identical (5) reduces as follows: For $q = 0$, $r = 0$, $A_i^0 = 1$, $B_i^0 = 1$, and $Q_{00} = n$; while in every other case $A_i^q B_i^r = 0$, $Q_{qr} = 0$. The summations with respect to q and r, therefore disappear.

Consider now the summations

$$\underset{j_1=1}{\overset{n}{S}}\ \underset{j_2=1}{\overset{n}{S}} \cdots \underset{j_k=1}{\overset{n}{S}}\ P^{j_1} P^{j_2} \ldots P^{j_k}\,.$$

Since all the populations are the same we may drop the j by actually carrying out the indicated summations. If, then, there are c repetitions among the k pairs of integers $\alpha_h\beta_h$, in which $\alpha_1\beta_1$, $\alpha_2\beta_2$, $\cdots$ $\alpha_c\beta_c$, are repeated l_1, l_2, $\cdots$ l_c times respectively, then we have;

$$\underset{j_1=1}{S} \cdots \underset{j_k=1}{S}\quad P^{j_h}_{\alpha_h\beta_h} = \frac{k\,!\,C_k^n *}{l_1\,!\,l_2\,!\cdots l_c\,!}\,\Pi\, P_{\alpha_h\beta_h}$$

We thus arrive at the following corollary: The mathematical expectation, $\bar{p}_{ab}$, of the product moment, p_{ab}, in samples of n from a single infinite population having any law of distribution is given by

$$n(-n)^{a+b}\bar{p}_{ab} = \underset{\alpha_h,\beta_h=0}{\overset{a,b}{S}} \frac{(a\,!)\,(b\,!)}{\Pi\alpha_h\,!\ \Pi\beta_h\,!}\ \underset{k=1}{\overset{t}{S}}\left[\underset{h=1}{\overset{n}{S}}\ (-n_1)^{\alpha_h+\beta_h} + n_t\right]\frac{k\,!\,C_k^n}{l_1\,!\cdots l_c\,!}\,\Pi\, P^*_{\alpha_h\beta_h}. \quad (5')$$

Note: In deriving these general formulae it was assumed that $n > t$. There is however, no loss in generality in this assumption. For, if $t > n$, we may suppose that, $x_{n+1} = x_{n+2} = \cdots = x_t = 0$, and hence $P_{\alpha\beta}^{n+1} = \cdots = P_{\mu\nu}^t = 0$, and thus the above reasoning is still valid.

5. **Formulae for** $\bar{p}_{41}, \bar{p}_{32}, \bar{p}_{51}, \bar{p}_{42}, \bar{p}_{33}$. Formulae for $\bar{p}_{ab}$ in which $a + b = 5, 6, 7, 8$ have been obtained. But for $(a + b) > 6$ these formulae become very long, and since these will be of no use in the subsequent work, only those of order 5 and 6 are given below.

$$\bar{p}_{41} = n^{-6}\,\{(n_1^5 - n_1)SP_{41}^i + 2nn_2^2 S(2P_{30}^i P_{11}^j + 3P_{21}^i P_{20}^j)\}$$

$$+ n^{-5}\,\{(n_1^4 + n_1)S(P_{40}^i B_i + 4P_{31}^i A_i) + 6nn_2 S(P_{20}^j B_i P_{20}^j + 2P_{11}^{i\dagger} A_i P_{20}^j)\}$$

* This is a generalization of Pepper's results for $N = \infty$. See *Biometrika* Vol. XXI, pp. 231–240.

† The symbol $P_{11}^i A_i P_{20}^j$ is an abbreviation of the full term $(A_i + A_j)\,(P_{11}^i P_{20}^j + P_{11}^j P_{20}^i)$. Similar abbreviations will be used in the other formulae.

$$+ 2n^{-4} \{(n_1^3 + 1) S(2P^i_{30}A_iB_i + 3P^i_{21}A^2_i) - 2Q_{11}SP^i_{30} - 3Q_{20}SP^i_{21}\}$$

$$+ 2n^{-3} \{(n_1^2 - 1)S(2P^i_{11}A^3_i 3P^i_{20}A^2_iB_i) + 2Q_{30}SP^i_{11} + 3Q_{21}SP^i_{20}\} + n^{-1}Q_{41}\,. \tag{6}$$

$$\bar{p}_{32} = n^{-6} \{(n_1^5 - n_1)SP^i_{32} + nn_2^2S(P^i_{30}P^j_{02} + 6P^i_{21}P^j_{11} + 3P^i_{20}P^j_{12})\}$$

$$+ n^{-5} \{(n_1^4 - 1)S(2P^i_{31}B_i + 3P^i_{22}A_i) + 3nn_2S(P^i_{20}P^j_{11}B_i + [P^i_{20}P^j_{02}$$

$$+ 4P^i_{11}P^j_{11}]\, A_i)\} + n^{-4} \{(n_1^3 + 1)S(P^i_{30}B^2_i + 6P^i_{21}A_iB_i$$

$$+ 3P^i_{12}A^2_i) - Q_{02}SP^i_{30} - 6Q_{11}SP^i_{21} - 3Q_{20}SP^i_{12}\} + n^{-3} \{(n_1^2 - 1)S(3P^i_{20}A_iB^2_i$$

$$+ 6P_{11}A_iB_i + P^i_{02}A^3_i) + 3Q_{12}SP^i_{20} + 6Q_{21}SP^i_{11} + Q_{30}SP^i_{02}\} + n^{-1}Q_{32}\,. \tag{7}$$

$$\bar{p}_{51} = n^{-7} \{(n_1^6 + n_1)SP^i_{51} + 5(n_1^4 + n_1^2 + n_2)S(P^i_{40}P^j_{11}$$

$$+ 2P^i_{31}P^j_{20}) - 10(2n_1^3 - n_2)SP^i_{21}P^j_{30} + 30(3n_1^2 + n_3)SP^i_{20}P^j_{20}P^k_{11}\} + n^{-6}\{(n_1^5$$

$$+ 1)S(P^i_{50}B_i + 5P^i_{41}A_i) + 10(n_1^3 + 1)S^*[2P^i_{30}P^j_{20}B_i + (2P^i_{30}P^j_{11}$$

$$+ 3P^i_{21}P^j_{20})A_i] - 10nn_2S^*[2P^i_{30}P^j_{20}B_j + (2P^i_{30}P^j_{11} + 3P^i_{21}P^j_{20})A_j]\}$$

$$+ 5n^{-5} \{(n_1^4 - 1)S(P^i_{40}A_iB_i + 2P^i_{31}A^2_i) + 6nn_2S(P^i_{20}P^j_{20}A_iB_i + 2P^i_{20}P^j_{11}A^2_i)$$

$$+ Q_{11}S(P^i_{40} + 6P^i_{20}P^j_{20}) + 2Q_{20}S(P^i_{31} + 6P^i_{20}P^j_{11})\} + 10n^{-4}\{(n_1^3$$

$$+ 1)S(P^i_{30}A^2_iB_i + P^i_{21}A^3_i) - Q_{21}SP^i_{30} - Q_{30}SP^i_{21}\} + 5n^{-3}\{(n_1^3 - 1)S(2P^i_{20}A^3_iB_i$$

$$+ P^i_{11}A^4_i) + 2Q_{31}SP^i_{20} + Q_{40}SP^i_{11}\} + n^{-1}Q_{51}\,. \tag{8}$$

$$\bar{p}_{42} = n^{-7}\{(n_1^6 + n_1)SP^i_{42} + (n_1^4 + n_1^2 + n_2)S(P^i_{10}P^j_{02} + 8P^i_{31}P^j_{11}$$

$$+ 6P^i_{22}P^j_{20}) + 4(2n_1^3 - n_2)S(P^i_{30}P^j_{12} + 3P^i_{21}P^j_{21}) + 6(3n_1^2 + n_3)S(P^i_{20}P^j_{20}P^k_{02}$$

$$+ 4P^i_{20}P^j_{11}P^k_{11})\} + 2n^{\ 6}\{(n_1^5 + 1)S(P^i_{41}B_i + 2P^i_{32}A_i) + 2(n_1^3 + 1)S[(2P^i_{30}P^j_{11}$$

$$+ 3P^i_{21}P^j_{20})B_i + (P^i_{30}P^j_{02} + 6P^i_{21}P^j_{11} + 3P^i_{12}P^j_{20})A_i] - 2nn_2S[(2P^i_{30}P^j_{11}$$

$$+ 3P^i_{21}P^j_{20})B_j + (P^i_{30}P^j_{12} + 6P^i_{21}P^j_{11} + 3P^i_{12}P^j_{20})A_j]\} + n^{-5}\{(n_1^4 - 1)S(P^i_{40}B^2_i$$

$$+ 8P^i_{31}A_iB_i + 6P^i_{22}A^2_i) + 6nn_2S[P^i_{20}P^j_{21}B^2_i + 4P^i_{20}P^j_{11}A_iB_i + (P^i_{20}P^j_{02}$$

$$+ 4P^i_{11}P^j_{11})A^2_i] + Q_{02}S(P^i_{40} + 6P^i_{20}P^j_{20}) + 8Q_{11}S(P^i_{31} + 3P^i_{20}P^j_{11})$$

$$+ 6P_{20}S(P^i_{22} + P^i_{20}P^j_{02} + 4P^i_{11}P^j_{11})\} + 4n^{-4}\{(n_1^3 + 1)S(P^i_{30}A_iB^2_i$$

$$+ 3P^i_{21}A^2_iB_i + P^i_{12}A^3_i) - Q_{12}SP^i_{30} - 3Q_{21}SP^i_{21} + Q^i_{30}SP^i_{12}\}$$

$$+ n^{-3}\{S(6P^i_{20}A^2_iB^2_i + 8P^i_{11}A^3_iB_i + P^i_{02}A^4_i) + Q_{40}SP^i_{02} + 8Q_{31}SP^i_{11}$$

$$+ 6Q_{22}SP^i_{20}\} + n^{-1}Q_{42}\,. \tag{9}$$

$$\bar{p}_{33} = n^{-7}\{(n_1^6 + n_1)SP^i_{33} + 3(n_1^4 + n_1^2 + n_2)S(P^i_{31}P^j_{p2} + 3P^i_{22}P^j_{11}$$

$$+ P^i_{13}P^j_{20}) - (2n_1^3 - n_2)\, S(P^i_{30}P^j_{03} + 9P^i_{21}P^j_{12}) + 9(3n_1^3 + n_3)S(P^i_{20}P^j_{11}P^k_{02}$$

* The repetition of this expression signifies that A and B factors are coupled only with those P factors which have corresponding indices.

$$+ 4P'_{11}P'_{11}P^k_{11})\} + 3n^{-6}\{(n_1^5 + 1)S(P'_{32}B_i + P'_{23}A_i) + (n_1^3 + 1)\, S[(P'_{30}P'_{02}$$
$$+ 6P'_{21}P'_{11} + 3P'_{12}P'_{20})B_i + (P'_{03}P'_{20} + 6P'_{12}P'_{11} + 3P'_{21}P'_{02})A_i]$$
$$- nn_2S[(P'_{30}P'_{02} + 6P'_{21}P'_{11} + 3P'_{12}P'_{20})B_j + (P'_{03}P'_{20} + 6P'_{12}P'_{11}$$
$$+ 3P'_{21}P'_{02})A_j]\} + 3n^{-5}\{(n_1^4 - 1)S(P'_{31}B_i^2 + 3P'_{22}A_iB_i + P'_{13}A_i^2)$$
$$+ 3n_1n_2S[P'_{20}P'_{11}B_i^2 + (P'_{20}P'_{02} + 4P'_{11}P'_{11})A_iB_i + P'_{11}P'_{02}A_i^2)]$$
$$+ S[Q_{02}(P'_{31} + 3P'_{20}P'_{11}) + 3Q_{11}(P'_{22} + P'_{20}P'_{02} + 4P'_{11}P'_{11}) + Q_{20}(P'_{13}$$
$$+ 3P'_{02}P'_{11})]\} + n^{-4}\{(n_1^3 + 1)S(P'_{30}B_i^3 + 9P'_{21}A_iB_i^2 + 9P'_{12}A_i^2A_i + P'_{03}A_i^3)$$
$$- S(Q_{03}P'_{30} + 9Q_{12}P'_{21} + 9Q_{21}P'_{12} + Q_{30}P'_{03})\} + 3n^{-3}\{(n_1^2 - 1)S(P'_{20}A_iB_i^3$$
$$+ 3P'_{11}A_i^2B_i^2 + P'_{02}A_i^3B_i) + S(Q_{13}P'_{20} + 3Q_{22}P'_{11} + Q_{31}P'_{02})\} + n^{-1}Q_{33}. \quad (10)$$

Chapter III. The Mathematical Expectation of the Variance of p_{ab}

1. The Symbols $_2m_{p_{ab}}$ and $_2M_{p_{ab}}$. Denoting the variance of p_{ab} by m and the mathematical expectation of $_2m_{p_{ab}}$ by $_2M_{p_{ab}}$, we have the definition,

$$_2m_{p_{ab}} = \{n^{-1}S(x_i - x)^a(y_i - y)^b - \bar{p}_{ab}\}^2$$
$$= n^{-2}S^2(x_i - x)^a(y_i - y)^b - 2n^{-1}\bar{p}_{ab}S(x_i - x)^a(y_i - y)^b + \bar{p}_{ab}^2, \text{ and}$$
$$_2M_{p_{ab}} = E(_2m_{p_{ab}}) = E\{n^{-2}S^2(x_i - x)^a(y_i - y)^b - 2n^{-1}\bar{p}_{ab}S(x_i - x)^a(y_i - y)^b + \bar{p}_{ab}^2\}$$
$$= n^{-2}E[S(x_i - x)^{2a}(y_i - y)^{2b}] + 2n^{-2}E[S(x_i - x)^a(x_j - x)^a(y_i - y)^b(y_j - y)^b]$$
$$- 2n^{-1}\bar{p}_{ab}E[S(x_i - x)^a(y_i - y)^b] + \bar{p}_{ab}^2 = n^{-1}\bar{p}_{2a2b}$$
$$+ 2n^{-2}E[S(x_i - x)^a(y_i - y)^b(x_j - x)^a(y_j - y)^b] - \bar{p}_{ab}^2. \quad (3.11)$$

Before attempting to expand the right hand side of (3.11) for any values a, b we shall derive the formula for $_2M_{p_{11}}$ to illustrate the procedure.

2. The Mathematical Expectation of $_2m_{p_{11}}$. By (3.11) we have

$$_2M_{p_{11}} = n^{-1}\bar{p}_{22} + 2n^{-2}E[S(x_i - x)(y_i - y)(x_j - x)(y_j - y)] - \bar{p}_{11}^2. \quad (3.21)$$

The first term is given by (4) and the last by (1). The only new term is the middle one. To expand it let us write it in terms of U and V. We then have:

$$n^{-2}SE[(x_i - x)(y_i - y)(x_j - x)(y_j - y_j)] = n^{-2}SE[(U_i + A_i)(V_i$$
$$+ B_i)(U_j + A_j)(V_j + B_j)] = n^{-2}\{SE[U_iV_iU_jV_j + (U_iV_iU_jB_j + U_jV_jU_iB_i)$$
$$+ (U_iV_iV_jA_j + U_jV_jV_iA_i) + (U_iV_iA_jB_j + U_jV_jA_iB_i)$$
$$+ (U_iV_jA_jB_i + U_jV_iA_iB_j) + U_iU_jB_iB_j + V_iV_jA_iA_j + 4 \text{ vanishing terms}$$
$$+ A_iB_iA_jB_j]\}. \quad (3.22)$$

The evaluation of the last term is very simple. For

$$SE(A_iB_iA_jB_j) = S(A_iB_iA_jB_j),$$

and from the elementary theory of symmetric functions we have:

$$S(A_iB_iA_jB_j) = \frac{S^2(A_iB_i) - S(A_i^2B_i^2)}{2}.$$

Hence

$$SE(A_iB_iA_jB_j) = \frac{S^2(A_iB_i) - S(A_i^2B_i^2)}{2} = \frac{Q_{11}^2 - Q_{22}}{2}. \qquad (3.23)$$

To expand the first term and also the remaining ones, we return to the u, v, notation defined in Chapter I. We then write

$$SE(U_iV_iU_jV_j) = n^{-4}SE[(n_1u_i - u_1 - \cdots)(n_1v_i - v_1 - \cdots)$$
$$(n_1u_j - u_1 - \cdots)(n_1v_j - v_1 - \cdots)].$$

The only terms which can appear in the expansion of the right hand side of the last equation have the following form:

$$E(u_i^2v_i^2), \qquad E(u_i^2v_j^2), \qquad E(u_iv_iu_jv_j),$$

i.e., exactly those which appear in the evaluation of $\bar{p}_{22}$. Remembering the symmetry, there will be no loss in generality if we take for i and j the integers 1 and 2. To find the coefficients of the three characterstic terms, the above summation may be broken up as follows:

$$n^4SE(U_iV_iU_jV_j) = E[(n_1u_1 - u_2 - \cdots)(n_1v_1 - v_2 - \cdots)(n_1u_2 - u_1 - \cdots)$$
$$(n_1v_2 - v_1 - \cdots)] + E\{[n_1u_1 - u_2 - \cdots)(n_1v_1 - v_2 - \cdots) + (n_1u_2 - u_1$$
$$- \cdots)(n_1v_2 - v_1 - \cdots)]\underset{3}{S}(n_1u_i - u_1 - \cdots)(n_1v_i - v_1 - \cdots)\} + \underset{3}{SE}[(n_1u_i$$
$$- u_1 - \cdots)(n_1v_i - v_1 - \cdots)(n_1u_j - u_1 - \cdots)(n_1v_j - v_1 - \cdots)]. \qquad (3.24)$$

Writing the three terms in a row and their coefficients from the three parts of (3.24) in columns below these terms, we get the following scheme:

	$E(u_1^2v_1^2)$	$E(u_1^2v_2^2 + u_2^2v_1^2)$	$E(u_1v_1u_2v_2)$
	n_1^2	n_1^2	$(n_1^2 + 1)^2$
	$n_2(n_1^2 + 1)$	$-2n_1n_2$	$2n_2^3$
	$\frac{n_2n_3}{2}$	$\frac{n_2n_3}{2}$	$2n_2n_3$
Total coeff.	$\frac{nn_1(2n_1 - 1)}{2}$	$\frac{-nn_3}{2}$	$n(n_1^3 + n_1^2 - 3n_1 + 3)$.

With the aid of the above equations we finally get:

$$SE(U_i V_i U_j V_j) = n^{-4}\left\{\frac{n_1 n(2n_1-1)}{2}\overset{*}{S}P^i_{22} - \frac{nn_3}{2}SP^i_{20}P^j_{02}\right.$$
$$\left. + n(nn_1^2 - 3n_2)SP^i_{11}P^j_{11}\right\}$$

Proceeding in the same way we find:

$$SE(U_i V_i U_j B_j + U_j V_j U_i B_i) = n^{-3}(2n_1^2 + n_2)SP^i_{21}B_i$$
$$SE(U_i V_i V_j A_j + U_j V_j V_i A_i) = n^{-3}(2n_1^2 + n_2)SP^i_{12}A_i$$
$$SE(U_i V_i A_j B_j + U_j V_j A_i B_i) = -nn_2 SP^i_{11}A_i B_i + (n_1^2 + n_2)Q_{11}SP^i_{11}$$
$$SE(U_i V_j A_j B_i + U_j V_i A_i B_j) = 2nSP^i_{11} - Q_{11}SP^i_{11}$$
$$SE(U_i U_j B_i B_j + V_i V_j A_i A_j) = nS(P^i_{20}B_i^2 + P^i_{02}A_i^2) - \tfrac{1}{2}S(Q_{20}P^i_{02} + Q_{02}P^i_{20}).$$

Collecting terms and simplifying we finally get:

$$_2M_{p_{11}} = n^{-4}\{n_1^2 SP^i_{22} + S(P^i_{20}P^j_{02} + 2P^i_{11}P^j_{11}) - n^2S(P^i_{11})^2\}$$
$$+ 2n^{-3}n_1\{S(P^i_{21}B_i + P^i_{12}A_i)\} + n^{-2}\{S(P^i_{20}B_i^2 + 2P^i_{11}A_iB_i + P^i_{02}A_i^2)\}. \quad (11)$$

Corollary 1. In case $X_i = Y_i$, i.e., when the set of populations are univariate, (11) becomes

$$_2M_{p_{20}} = n^{-4}\{n_1^2 S[P^i_{40} - (P^i_{20})^2] + 4SP^i_{20}P^j_{20}\} + 4n^{-3}n_1 SP^i_{30}A_i + 4n^{-2}SP^i_{20}A_i^2. \quad (11')$$

This is Tchouproff's formula for the expected value of the variance of samples of n.[10]

Corollary 2. In case the n populations are identical (11) becomes

$$_2M_{p_{11}} = n^{-3}n_1[n_1P_{22} + P_{20}P_{02} - n_2P_{11}^2]. \quad (11'')^{11}$$

3. The Mathematical Expectation of $_2M_{p_{ab}}$. We now return to the general equation

$$_2M_{p_{ab}} = n^{-1}\bar{p}_{2a2b} - \bar{p}^2_{ab} + 2n^{-2}\overset{n}{\underset{i=1,j=1}{S}} E(x_i - x)^a(y_i - y)^b(x_j - x)^a(y_j - y)^b. \quad (3.11)$$

* Since $E(u_i^2v_i^2) = P^i_{22}$, $E(u_i^2v_j^2) = P^i_{20}P^j_{02}$, etc.

[10] See *Biometrika*, Vol. XIII p. 295.

[11] See *Biometrika*, Vol. XXI p. 234, Cor. 1.

The first two terms are given by (5). To evaluate the last term we write:

$$SE[(x_i - x)^a(y_i - y)^b(x_j - x)^a(y_j - y)^b] = SE[(U_i + A_i)^a(V_i + B_i)^b(U_j + A_j)^a$$

$$(V_j + B_j)^b] = SE(U_i^a V_i^b U_j^a V_j^b) + \overset{a,a,b,b,}{\underset{r_1,r_2,r_3,r_4=0}{S}} C_{r_1}^a C_{r_2}^a C_{r_3}^b C_{r_4}^b$$

$$\overset{n}{\underset{1}{S}} E(U_i^\alpha V_i^\gamma U_j^\beta V_j^\delta A_i^{r_1} B_i^{r_3} A_j^{r_2} B_j^{r_4}) = n^{-2(a+b)} SE\{(n_1u_i - \cdots)^a(n_1v_i - \cdots)^b$$

$$(n_1u_j - \cdots)^a(n_1v_j - \cdots)^b + \underset{r_1 \quad r_4}{S} \; n^{(r_1+r_2+r_3+r_4)} C_{r_1}^a \cdots C_{r_4}^b SE[(n_1u_i - \cdots)^a$$

$$(n_1v_i - \cdots)^\gamma(n_1u_j - \cdots)^\beta(n_1v_j - \cdots)^\delta A_i^{r_1}B_i^{r_3}A_j^{r_2}B_j^{r_4}, \tag{3.31}$$

where $\alpha = a - r_1$, $\beta = a - r_2$, $\gamma = b - r_3$, $\delta = b - r_4$.

The right hand side of (3.31) has been broken up into two parts because the first part is symmetrical, while the second part, in general, is not except when $r_1 = r_2$, and $r_3 = r_4$.

Let us now consider the expression

$$SE[(n_1u_i - \cdots)^a(n_1v_i - \cdots)^b(n_1u_j - \cdots)^a(n_1v_j - \cdots)^b]. \tag{3.32}$$

This is a double summation in which $c_{ij} = c_{ji}$, and in which the diagonal terms, c_{ii}, are missing.

Consider next a general term of k factors from the expansion of each bracket of (3.32). As we are dealing with symmetric functions, there will be no loss in generality if we consider the first k subscripts only; and if we let the lower limits of the exponents of the u's and v's begin with zero we may consider that each parenthesis of a given bracket contributes exactly k factors. Such a term, omitting the coefficient, may be written as follows:

$$E(u_1^{\alpha_1} \cdots u_k^{\alpha_k} v_1^{\beta_1} \cdots v_k^{\beta_k} u_1^{\alpha_1'} \cdots u_k^{\alpha_k'} v_1^{\beta_1'} \cdots v_k^{\beta_k'}) = \prod^{k} E(u_h^{\alpha_h+\alpha_h'} v_h^{\beta_h+\beta_h'})$$

$$= \prod_{h=1}^{k} P^h(\alpha_h + \alpha_h')\,(\beta_h + \beta_h'). \tag{3.33}$$

This term occurs in every one of the $\frac{1}{2}nn_1$ brackets of (3.32), having the same numerical coefficient in every one of them, which is

$$\frac{(a\,!)^2\,(b\,!)^2}{\Pi\alpha_h\,!\,\Pi\alpha_h'\,!\,\Pi\beta_h\,!\,\Pi\beta_h'\,!}. \tag{3.34}$$

To obtain the n_1 coefficient of (3.33) we break up (3.32) into the following partial summations:

$$E[(n_1u_i - \cdots)^a(n_1v_i - \cdots)^b(n_1u_j - \cdots)^a(n_1v_j - \cdots)^b] = E[(n_1u_1 - \cdots)^a$$

$$(n_1v_1 - \cdots)^b\,(n_1u_2 - \cdots)^a(n_1v_2 - \cdots)^b] + \cdots + E[(n_1u_{k-1} - \cdots)^a$$

$$(n_1 v_{k-1} - \cdots)^b (n_1 u_k - \cdots)^a (n_1 v_k - \cdots)^b] + \mathop{S}_{i=1}^{k} E\Big[(n_1 u_i - \cdots)^a \cdot$$

$$(n_1 v_i - \cdots)^b \mathop{S}_{j=k+1}^{n} (n_1 u_j - \cdots)^a (n_1 v_j - \cdots)^b\Big] + \mathop{S}_{i,j=k+1}^{n} [E\{(n_1 u_i - \cdots)^a$$

$$(n_1 v_i - \cdots)^b (n_1 u_j - \cdots)^a (n_1 v_j - \cdots)^b\}] .$$

From this equation we get for the total coefficient in n of the term (3.33) the following expression:

$$\mathop{S}_{h,h'=1}^{k} (-n_1)^{\alpha_h + \alpha'_{h'} + \beta_h + \beta'_{h'}} + n_k \mathop{S}_{h=1}^{k} [(-n_1)^{\alpha_h + \beta_h} + (-n_1)^{\alpha'_h + \beta'_h}] + C_2^{n_k} .$$

The following restrictions on the α's and β's must be observed.

$$\text{(a)} \begin{array}{l} \alpha_1 + \alpha_2 + \cdots + \alpha_k = a \\ \alpha'_1 + \alpha'_2 + \cdots + \alpha'_k = a \end{array} \qquad \text{(b)} \begin{array}{l} \beta_1 + \beta_2 + \cdots + \beta_k = b \\ \beta'_1 + \beta'_2 + \cdots + \beta'_k = b \end{array}$$

$$\text{(c)} \quad \alpha_h + \alpha'_h + \beta_h + \beta'_h \neq 1 .$$

From (c) we obtain the upper limit of k, namely: $t = a + b$.

Combining the various above equations we finally obtain:

$$(n)^{2(a+b)} S(U_i^a V_i^b U_j^a V_j^b) = (a\,!)^2 (b\,!)^2 \mathop{S}_{j_h=1}^{n} \mathop{S}_{\alpha_h, \alpha'_h, \beta_h, \beta'_h = 0}^{n,b}$$

$$\mathop{S}_{k=1}^{t} \left\{ \mathop{S}_{h,h'=1}^{k} (-n_1)^{\alpha_h + \beta_h + \alpha'_{h'} + \beta'_{h'}} + n_k \mathop{S}_{h=1}^{k} [(-n_1)^{\alpha_h + \beta_h} + (-n_1)^{\alpha'_h + \beta'_h}] + C_2^{n_k} \right\}$$

$$\frac{\Pi P^{J_h}_{(\alpha_h + \alpha'_h)(\beta_h + \beta'_h)}}{\Pi \alpha_h\,!\ \Pi \alpha'_h\,!\ \Pi \beta_h\,!\ \Pi \beta'_h\,!} . \tag{3.35}$$

Turning to the second part of (3.31) let us consider the expression

$$\mathop{S}_{i=1,j=1}^{n} E[(n_1 u_i - \cdots)(n_1 v_i - \cdots)(n_1 u_j - \cdots)(n_1 v_j - \cdots) A_i^{r_1} B_i^{r_3} A_j^{r_2} B_j^{r_4}$$

for a given set of r's. The term (3.33) may also be considered as a general term of this last expression; of course, the exponents of the u's and v's will be different in this case. In order to evaluate the complete coefficient of a term like (3.33) we again write;

$$SE[(n_1 u_i - \cdots)^\alpha (n_1 v_i - \cdots)^\gamma (n_1 u_j - \cdots)^\beta (n_1 v_j - \cdots)^\delta A_i^{r_1} B_i^{r_3} A_j^{r_2} B_j^{r_4}$$

$$= E[(n_1 u_1 - \cdots)^\alpha (n_1 v_1 - \cdots)^\gamma (n_1 u_2 - \cdots)^\beta (n_1 v_2 - \cdots)^\delta A_1^{r_1} B_1^{r_3} A_2^{r_2} B_2^{r_4}]$$

$$+ E[n_1 u_2 - \cdots)^\alpha (n_1 v_2 - \cdots)^\gamma (n_1 u_1 - \cdots)^\beta (n_1 v_1 - \cdots)^\delta A_2^{r_1} B_2^{r_3} A_1^{r_2} B_1^{r_4}$$

$$+ \cdots + E[(n_1 u_k - \cdots)^\alpha (n_1 v_k - \cdots)^\gamma (n_1 u_{k-1} - \cdots)^\beta (n_1 v_{k-1} - \cdots)^\delta$$

$$A_k^{r_1}B_k^{r_3}A_{k-1}^{r_2}\ B_{k-1}^{r_4}] + \overset{k}{\underset{i=1}{S}} E[(n_1u_i - \cdots)^\alpha(n_1v_i - \cdots)^\gamma A_i^{r_1}B_i^{r_3} \overset{n}{\underset{j=k+1}{S}}$$

$$(n_1u_j - \cdots)^\beta(n_1v_j - \cdots)^\delta A_j^{r_2}B_j^{r_4}] + \overset{k}{\underset{j=1}{S}} E[(n_1u_j - \cdots)^\beta(n_1v_j - \cdots)^\delta$$

$$A_j^{r_2}B_j^{r_4} \overset{n}{\underset{i=k+1}{S}} (n_1u_i - \cdots)^\alpha(n_1v_i - \cdots)^\gamma A_i^{r_1}B_i^{r_3}] + \overset{n}{\underset{i,j=k+1}{S}} E[(n_1u_i - \cdots)^\alpha$$

$$n_1v_i - \cdots)^\gamma(n_1u_j - \cdots)^\beta(n_1v_j - \cdots)^\delta A_i^{r_1}B_i^{r_3}A_j^{r_2}B_j^{r_4}]. \tag{3.36}$$

It is now quite easy to write down the complete coefficient of a term of the form (3.33). The numerical coefficient of this term is the same in every bracket of (3.36), and is

$$\frac{(-1)\overset{4}{\underset{1}{S}}{}_{r_i}(a-r_1)!\,(a-r_2)!\,(b-r_3)!\,(b-r_4)!}{\Pi\alpha_h!\,\Pi\alpha_h'!\,\Pi\beta_h!\,\Pi\beta_h'!}. \tag{3.37}$$

The coefficient in n_1 and $A_i^{r_1}\,B_i^{r_3}\,A_j^{r_2}\,B_j^{r_4}$ is broken up by (3.36) into the following four parts:

I. $\overset{k}{\underset{h=1,h'=1}{S}} (-n_1)^{\alpha_h+\alpha_{h'}'+\beta_h+\beta_{h'}'} A_h^{r_1}B_h^{r_3}A_{h'}^{r_2}B_{h'}^{r_4}$, from the first $k(k-1)$ brackets.

II. $$\overset{k}{\underset{h=1}{S}} (-n_1)^{\alpha_h+\beta_h} A_h^{r_1}B_h^{r_3} \overset{n}{\underset{h'=k+1}{S}} A_{h'}^{r_2}B_{h'}^{r_4} = \overset{k}{\underset{h=1}{S}} (-n_1)^{\alpha_h+\beta_h} A_h^{r_1}B_h^{r_3} \left[Q_{r_2r_4} - \overset{k}{\underset{h'=1}{S}} A_{h'}^{r_2}B_{h'}^{r_4}\right],$$

from the next $k(n-k)$ brackets. Similarly

III. $\overset{k}{\underset{h'=1}{S}} (-n_1)^{\alpha_{h'}'+\beta_{h'}'} A_{h'}^{r_2}B_{h'}^{r_4}\left[Q_{r_1r_3} - \overset{k}{\underset{h=1}{S}} A_h^{r_1}B_h^{r_3}\right]$, from the next $k(n-k)$.

And finally:

IV. $$\overset{n}{\underset{i,j=k+1}{S}} A_i^{r_1}B_i^{r_3}A_j^{r_2}B_j^{r_4} = \overset{n}{\underset{1}{S}}A_h^{r_1}B_h^{r_3}\overset{n}{\underset{1}{S}}A_h^{r_2}B_h^{r_4} - \overset{n}{\underset{1}{S}}A_h^{(r_1+r_2)}B_h^{(r_3+r_4)}$$

$$- \overset{k}{S}\ A_h^{r_1}B_h^{r_3}A_h^{r_2}B_h^{r_4} - \overset{n}{\underset{h=1}{S}} A_h^{r_1}B_h^{r_3} \overset{k}{\underset{h'=1}{S}} A_{h'}^{r_2}B_{h'}^{r_4} - \overset{n}{\underset{h=1}{S}} A_h^{r_2}B_h^{r_4} \overset{k}{\underset{h=1}{S}} A_h^{r_1}B_h^{r_3}$$

$$+ 2 \overset{}{\underset{h=1}{S}} A_h^{r_1}B_h^{r_3} \overset{k}{\underset{h'=1}{S}} A_{h'}^{r_2}B_{h'}^{r_4} = Q_{r_1r_3}Q_{r_2r_4} - Q_{(r_1+r_2)(r_3+r_4)} - Q_{r_1r_3}\overset{k}{\underset{1}{S}}A_h^{r_2}B_h^{r_4}$$

$$- Q_{r_2r_4}\overset{k}{\underset{1}{S}}A_h^{r_1}B_h^{r_3} - \overset{k}{\underset{h,h'=1}{S}} A_h^{r_1}B_h^{r_3}A_{h'}^{r_2}B_{h'}^{r_4} + 2\overset{k}{\underset{h=1}{S}} A_h^{r_1}B_h^{r_3} \overset{k}{\underset{h'=1}{S}} A_{h'}^{r_2}B_{h'}^{r_4},$$ from the last c_2^{n-k} brackets.

The restrictions on the α's and β's differ from those given above in that a is replaced by $a - r_1$ and $a - r_2$, and b by $b - r_3$ and $b - r_4$; and from the restriction (c) we get for the upper limit of k, in this case,

$$t_1 = \frac{\alpha + \beta + \gamma + \delta}{2} = a + b - \frac{r_1 + r_2 + r_3 + r_4}{2}$$

when $\overset{4}{\underset{1}{S}} r_\iota$ is even, or the greatest integer less then $\frac{S\alpha}{2}$ when Sr_ι is odd.

Combining (3.37) with $C^a_{r_1} \cdots C^b_{r_4}$ we get for the general numerical coefficient in the expansion of the second part of (3.31), the expression

$$\frac{(-1)^{Sr_\iota}(a!)^2(b!)^2}{\Pi r_\iota! \, \Pi\alpha_h! \, \Pi\alpha'_h! \, \Pi\beta_h! \, \Pi\beta'_h!}.$$

By an obvious manipulation we have

$$\begin{aligned} \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} = & \underset{h,h'=1}{\overset{k}{S}} \left[(-n_1)^{\alpha_h+\beta_h+\alpha'_{h'}+\beta'_{h'}} - 1\right] A_h^{r_1} B_h^{r_3} A_{h'}^{r_2} B_{h'}^{r_4} + Q_{r_2r_4} \\ & \underset{h=1}{\overset{k}{S}} \left[(-n_1)^{\alpha_h+\beta_h} - 1\right] A_h^{r_1} B_h^{r_3} + Q_{r_1r_3} \underset{h=1}{\overset{k}{S}} \left[(-n_1)^{\alpha'_h+\beta'_h} - 1\right] A_h^{r_2} B_h^{r_4} \\ & - \underset{h=1}{\overset{k}{S}} A_h^{r_2} B_h^{r_4} \underset{h=1}{\overset{k}{S}} \left[(-n_1)^{\alpha_h+\beta_h} - 1\right] A_h^{r_1} B_h^{r_3} - \underset{h=1}{\overset{k}{S}} A_h^{r_1} B_h^{r_3} \\ & \underset{h=1}{\overset{k}{S}} \left[(-n_1)^{\alpha'_h+\beta'_h} - 1\right] A_h^{r_2} B_h^{r_4} + Q_{r_1r_3} Q_{r_2r_4} - Q_{(r_1+r_2)(r_3+r_4)}. \end{aligned} \tag{3.38}$$

Finally, combining the various equations we get the formula:

$$\begin{aligned} {}_2M_{p_{ab}} = {} & n^{-1}\bar{p}_{2a2b} - \bar{p}^2_{a\,b} + 2(n)^{-2(a+b+1)}(a\,!)^2(b\,!)^2 \underset{j\,h=1}{\overset{n}{S}} \underset{\alpha_h,\alpha'_{h'},\beta_h,\beta'_{h'}=0}{\overset{a,b}{S}} \underset{k=1}{\overset{t}{S}} \\ & \underset{h,h'=1}{\overset{k}{S}} (-n_1)^{\alpha_h+\beta_h+\alpha'_{h'}+\beta'_{h'}} + n_k \underset{h=1}{\overset{k}{S}} \left[(-n_1)^{\alpha_h+\beta_h} + (-n_1)^{\alpha'_h+\beta'_h}\right] + C_2^{n_k} \\ & \frac{\Pi P^{jh}(\alpha_h + \alpha'_h)(\beta_h + \beta'_h)}{\Pi\alpha_h\,!\,\Pi\beta_h\,!\,\Pi\alpha'_h\,!\,\Pi\beta'_h\,!} + 2(n)^{-2(a+b+1)}(a\,!)^2(b\,!)^2 \underset{j\,h=1}{\overset{n}{S}} \underset{r_1,r_2,r_3,r_4=0}{\overset{a,b}{S}} \frac{(-n)^{Sr_1}Sr_\iota}{\Pi r_1\,!} \\ & \underset{\alpha_h,\alpha'_{h'},\beta_h,\beta'_{h'}=0}{\overset{\alpha,\beta,\gamma,\delta}{S}} \underset{k=1}{\overset{t}{S}} \left\{ \underset{h,h'=1}{\overset{k}{S}} \left[(-n_1)^{\alpha_h+\beta_h+\alpha'_{h'}+\beta'_{h'}} - 1\right] A_h^{r_1} B_h^{r_3} A_{h'}^{r_2} B_{h'}^{r_4} \right. \\ & - \underset{h=1}{\overset{k}{S}} \left[(-n_1)^{\alpha_h+\beta_h} - 1\right] A_h^{r_1} B_h^{r_3} \underset{h=1}{\overset{k}{S}} A_h^{r_2} B_h^{r_4} - \underset{h=1}{\overset{k}{S}} \left[(-n_1)^{\alpha'_h+\beta'_h} - 1\right] \\ & A_h^{r_2} B_h^{r_4} \underset{h=1}{\overset{k}{S}} A_h^{r_1} B_h^{r_3} + Q_{r_2r_4} \underset{h=1}{\overset{k}{S}} \left[(-n_1)^{\alpha_h+\beta_h} - 1\right] A_h^{r_1} B_h^{r_3} \end{aligned}$$

$$+ Q_{r_1r_3} \mathop{S}_{h=1}^{k} [(-n_1)^{\alpha'_h+\beta'_h} - 1] A_h^{r_2} B_h^{r_4} + Q_{r_1r_3}Q_{r_2r_4} - Q_{(r_1+r_2)(r_3+r_4)}$$

$$\frac{\Pi P^{j_h}_{(\alpha+\alpha'_h)(\beta_h+\beta'_h)}}{\Pi\alpha_h!\,\Pi\beta_h!\,\Pi\alpha'_h!\,\Pi\beta'_h!}. \tag{12}$$

In case the n populations are identical the second part of (12) must vanish, and in the first part the summations

$$\mathop{S}_{j_h=1}^{n} \prod_{h=1}^{k} P^{j_h}_{(\alpha_h+\alpha'_h)(\beta_h+\beta'_h)} = \frac{k!\,C_k^n\,\Pi\,P_{(\alpha_h+\alpha'_h)(\beta_h+\beta'_h)}}{l_1!\,l_2!\cdots l_c!},$$

where $l_1, l_2, \cdots l_c$ are the number of repetitions of the pairs of integers $(\alpha_1 + \alpha'_1)(\beta_1 + \beta'_1), \cdots (\alpha_k + \alpha'_k)(\beta_k + \beta'_k)$, respectively.
We then have the following
Corollary: The mathematical expectation of the variance, ${}_2m_{p_{ab}}$, of the product moment, p_{ab}, in samples of n from a single infinite population is given by

$${}_2M_{p_{ab}} = \bar{p}_{2a2b} - \bar{p}^2_{ab} + 2(n)^{-2(a+b+1)}(a!)^2(b!)^2 \quad \mathop{S}^{a,a,b,b}$$

$$\mathop{S}_{k=1}^{t} \frac{k!\,C^{nk}}{l_1!\,l_2!\cdots l_c!}\left\{\mathop{S}_{h,h'=1}^{k} (-n_1)^{\alpha_h+\beta_h+\alpha'_{h'}+\beta'_{h'}} + n_k \mathop{S}_{h=1} [(-n_1)^{\alpha_h+\beta_h}\right.$$

$$\left. + (-n_1)^{\alpha'_h+\beta'_h}] + C_2^{nk}\right\} \frac{\prod_1^k P(\alpha_h + \alpha'_h)(\beta_h + \beta'_h)}{\Pi\alpha_h!\,\Pi\beta_h!\,\Pi\alpha'_h!\,\Pi\beta'_h!}. \tag{12'}$$

4. **The Formula for ${}_2M_{p_{21}}$.** Formula (12) can by no means be used mechanically. It does, however, summarize to a great extent the details in finding ${}_2M_{p_{ab}}$ for any given values a, b. Formulae for ${}_2M_{p_{21}}$, ${}_2M_{p_{31}}$ have been obtained, but the one for ${}_2M_{p_{31}}$ is too long to be included in the paper, especially since with a little work it can be easily derived by applying (12). The one for ${}_2M_{p_{21}}$ is given immediately below.

$${}_2M_{p_{21}} = n^{-6}\{n_1^2n_2^2S[P^i_{42} - (P^i_{21})^2] + n_2^2S[P^i_{40}P^i_{02} + 4(P^i_{30}P^j_{12} - n_2P^i_{31}P^j_{11})]$$

$$- 2n_2^2n_3SP^i_{22}P^j_{20} + (n_2^2 + 2)S(P^i_{20}P^j_{20}P^k_{02} + 8P^i_{20}P^j_{11}P^k_{11}) + 6SP^i_{20}P^j_{20}P^k_{02}\}$$

$$+ 2n^{-5}\{n_1n_2^2S(P^i_{41}B_i + 2P^i_{32}A_i - P^i_{30}P^i_{11}B_i - 2P^i_{21}P^i_{11}A_i)$$

$$- 4n_2n_4S(P^i_{30}B_iP^j_{11} + P^i_{12}A_iP^j_{20}) - 2n_2S[n_5P^i_{21}B_iP^j_{20} + 2(2n_2 - 3)P^i_{21}A_iP^j_{11}]$$

$$+ 6nSP^i_{21}P^j_{11}A_j + 4n_2S(P^i_{30}P^j_{11}B_j + P^i_{30}P^j_{02}A_j + P^i_{30}A_iP^j_{02} + 2P^j_{21}P^j_{20}B_j$$

$$+ P^i_{12}P^j_{20}A_j)\} + n^{-4}\{n_2^2S[P^i_{40}B_i^2 - (P^i_{20}B_i)^2] + 4SP^i_{20}P^j_{20}(B_i + B_j)^2$$

$$+ 3(n_2^2 + n_2)SP^i_{22}A_i^2 + 4SP^i_{20}P^j_{02}(A_i + A_j)^2 - 2n_4S[P^i_{20}P^j_{02}A_j^2$$

$$
\begin{aligned}
&+ 2P^{i}_{11}P^{j}_{11}(A_i + A_j)^2] + 16SP^{i}_{11}A_iP^{j}_{11}A_j - 4n_2^2S(P^{i}_{11}A_i)^2 \\
&+ 4(2n_2^2 + n_2)SP^{i}_{31}A_iB_i - 4n_4SP^{i}_{11}A_iB_iP^{j}_{20} - 8n_3SP_{11}P^{j}_{20}A_jB_j \\
&+ 8S(P^{i}_{11}B_iP^{i}_{20}A_j + P^{i}_{11}A_iP^{j}_{20}B_j) - 4n_2^2SP^{i}_{11}A_iP^{i}_{20}B_i \\
&- 2n_1n_2n^{-1}S(Q_{20}P^{i}_{22} + 2Q_{11}P^{i}_{31}) + 2n_2n^{-1}S[6Q_{11}P^{i}_{11}P^{j}_{20} \\
&+ Q_{20}(P^{i}_{20}P^{j}_{02} + 4P^{i}_{11}P^{j}_{11})]\} + 2n^{-4}\{nn_2S(2P^{i}_{30}A_iB_i^2 + 2P^{i}_{12}A_i^3 + 5P^{i}_{21}A_i^2B_i) \\
&- n_1S[Q_{20}(P^{i}_{21}B_i + P^{i}_{12}A_i) + 2Q_{11}8P^{i}_{30}B_i + 2P^{i}_{21}A_i)]\} + n^{-4}\{n^2S[P^{i}_{02}A_i^4 \\
&+ 4(P^{i}_{20}A_i^2B_i^2 + P^{i}_{11}A_i^2B_i)] - 2nS[(Q_{20}A_iB_i + Q_{11}A_i^2)P^{i}_{11} + Q_{11}P^{i}_{20}A_iB_i \\
&+ Q_{20}P^{i}_{02}A_i^2] + S[Q_{20}^2P^{i}_{02} + 4Q_{20}(Q_{11}P^{i}_{11} + Q_{11}^2P^{i}_{20})]\} . \qquad (13)[12]
\end{aligned}
$$

Chapter IV. The Mathematical Expectation of the Third Moment of p_{11}

1. The Mathematical Expectation of ${}_3m_{p_{11}}$. Following the notation of the last chapter we shall denote the third moment of p_{11} about its mean by ${}_3m_{p_{11}}$ and the mathematical expectation of ${}_3m_{p_{11}}$ by ${}_3M_{p_{11}}$. We have then by definition.

$$ {}_3m_{p_{11}} = \{n^{-1}S(x_i - x)(y_i - y) - \bar{p}_{11}\}^3 , $$

and by a well known formula we have:

$$ {}_3M_{p_{11}} = \overline{p^3_{11}} - 3_2M_{p_{11}}\bar{p}_{11} - \bar{p}^3_{11} . \qquad (4.11) $$

The last two terms of (4.11) are given by (1) and (11). To evaluate $\overline{p^3_{11}}$ we write:

$$
\begin{aligned}
\overline{\bar{p}^3_{11}} = E\{n^{-1}S(x_i - x)(y_i - y)\}^3 &= n^{-3}SE(x_i - x)^3(y_i - y)^3 \\
&+ 3n^{-3}SE(x_i - x)^2(y_i - y)^2(x_j - x)(y_j - y) \\
&+ 6n^{-3}SE(x_i - x)(y_i - y)(x_j - x)(y_j - y)(x_k - x)(y_k - y) .
\end{aligned}
$$

The first term is simply $n^{-2}\bar{p}_{33}$ which is given by (10). The evaluation of the second term is not essentially different from the evaluation of the left hand side of (3.22), and since all details have been given there we shall omit them here.

To evaluate the last expression let us write:

$$
\begin{aligned}
&SE(x_i - x)(y_i - y)(x_j - x)(y_j - y)(x_k - x)(y_k - y) \\
&= SE[(U_i + A_i)(V_i + B_i)(U_j + A_j)(V_j + B_j)(U_k + A_k)(V_k + B_k)] \\
&= SE(U_iV_iU_jV_jU_kV_k) + SE(U_iV_iU_jV_jU_kB_k) + \cdots + SE(A_iB_iA_jB_jA_kB_k) .
\end{aligned}
\qquad (4.12)
$$

[12] In case the n populations are identical this reduces to one of Pepper's formulae, *Biometrika*, Vol. XXI, p. 238, Cor. 1.

As there is a great deal of similarity among the various terms of the right hand side of (4.12), it will not be necessary to go into the details of the expansion of every one of them. We shall, therefore, indicate the details for the expansion of only two of them—one symmetrical and one non-symmetrical; and as the first two terms are of that type we shall use these for the purpose of illustration.

Using the u, v notation we have

$$SE(U_iV_iU_jV_jU_kV_k) = n^{-6}SE[(n_1u_i - \cdots)(n_1v_i - \cdots)(n_1u_j - \cdots)$$
$$(n_1v_j - \cdots)(n_1u_k - \cdots)(n_1u_k - \cdots)].$$

The maximum number of subscripts appearing in any term evidently being 3, we can write without any loss in generality:

$$SE[(n_1u_i - \cdots) \cdots (n_1v_k - \cdots)] = E[(n_1u_1 - \cdots)(n_1v_1 - \cdots)(n_1u_2 - \cdots)$$
$$n_1v_2 - \cdots)(n_1u_3 - \cdots)(n_1v_3 - \cdots)] + E\{(n_1u_1 - \cdots)(n_1v_1 - \cdots)[(n_1u_2 - \cdots)$$
$$(n_1v_2 - \cdots) + (n_1u_3 - \cdots)(n_1v_3 - \cdots)] + (n_1u_2 - \cdots)(n_1v_2 - \cdots)$$
$$(n_1u_3 - \cdots)(n_1v_3 - \cdots)\}\underset{4}{S}(n_1u_i - \cdots)(n_1v_i - \cdots) + E\{(n_1u_1 - \cdots)$$
$$(n_1v_1 - \cdots) + (n_1u_2 - \cdots)(n_1v_2 - \cdots) + n_1u_3 - \cdots)(n_1v_3 - \cdots)\}$$
$$\underset{4}{S}(n_1u_i - \cdots)(n_1v_i - \cdots)(n_1u_j - \cdots)(n_1v_j - \cdots) + \underset{4}{S}E\{(n_1u_i - \cdots)\cdots$$
$$(n_1v_k - \cdots)\}. \tag{4.13}$$

The coefficients of the various terms arising in this expansion can now be found quite easily. For example, the coefficient of $P^1_{3\,3}$, which is, of course, the same as the coefficient of $P^i_{3\,3}$, is easily found to be

$$n_1^2 + n_3(2n_1^2 + 1) + \frac{n_3n_4(n_1^2 + 2)}{2} + \frac{n_3n_4n_5}{6} = \frac{nn_1n_2(3n_1 - 2)}{6}.$$

To evaluate the summation $SE(U_iV_iU_jV_jU_kB_k) = n^{-5}SE[(n_1u_i - \cdots)$ $(n_1v_i - \cdots)$ $(n_1u_j - \cdots)(n_1v_j - \cdots)(n_1v_k - \cdots)B_k]$, we break it up into partial summations as follows:

$$SE[(n_1u_i - \cdots)(n_1v_i - \cdots)(n_1u_j - \cdots)(n_1v_j - \cdots)(n_1u_k - \cdots)B_k]$$
$$= E\{(n_1u_1 - \cdots)(n_1v_1 - \cdots)[(n_1u_2 - \cdots)(n_1v_2 - \cdots)(n_1u_3 - \cdots)B_3$$
$$+ (n_1u_2 - \cdots)B_2(n_1u_3 - \cdots)(n_1v_3 - \cdots)] + (n_1u_1 - \cdots)B_1(n_1u_2 - \cdots)$$
$$(n_1v_2 - \cdots)(n_1u_3 - \cdots)(n_1v_3 - \cdots)\} + E\{(n_1u_1 - \cdots)(n_1v_1 - \cdots)$$
$$[(n_1u_2 - \cdots)(n_1v_2 - \cdots) + (n_1u_3 - \cdots)(n_1v_3 - \cdots)] + (n_1u_2 - \cdots)$$
$$(n_1v_2 - \cdots)(n_1u_3 - \cdots)(n_1v_3 - \cdots)\}\underset{4}{S}(n_1u_j - \cdots)B_j + E\{(n_1u_1 - \cdots)$$
$$(n_1v_1 - \cdots)[(n_1u_2 - \cdots)B_2 + (n_1u_3 - \cdots)B_3] + (n_1u_2 - \cdots)(n_1v_2 - \cdots)$$

$$[(n_1u_1 - \cdots)B_1 + (n_1u_3 - \cdots)B_3] + (n_1u_3 - \cdots)(n_1v_3 - \cdots)$$

$$[(n_1u_1 - \cdots)B_1 + (n_1v_2 - \cdots)B_2]\}\underset{4}{S}(n_1u_j - \cdots)(n_1v_j - \cdots)$$

$$+ E\{(n_1u_1 - \cdots)(n_1v_1 - \cdots) + (n_1u_2 - \cdots)(n_1v_2 - \cdots) + (n_1u_3 - \cdots)$$

$$(n_1v_3 - \cdots)\}\ \underset{4}{S}(n_1u_i - \cdots)(n_1v_i - \cdots)(n_1u_j - \cdots)B_j + E\{(n_1u_1 - \cdots)B_1$$

$$+ (n_2u_2 - \cdots)B_2 + (n_1u_3 - \cdots)B_3\}\underset{4}{S}(n_1u_i - \cdots)(n_1v_i - \cdots)$$

$$(n_1u_j - \cdots)(n_1v_j - \cdots) + E\underset{4}{S}(n_1u_i - \cdots)(n_1v_i - \cdots)(n_1u_j - \cdots)$$

$$(n_1v_j - \cdots)(n_1u_k - \cdots)B_k. \qquad (4.14)$$

The expansion of (4.14) is not as difficult as it appears for only two subscripts can appear in any term: the explicit appearance of the subscript 3 is due to the fact that we are dealing with a triple summation. We, consequently, do not need to expand those parentheses in which B appears.

We shall now, without any further details, state the final result, which is:

$$\begin{aligned}
{}_3M_{p_{11}} = {} & n^{-6}\{S[n_1^3P^i_{33} - P^i_{30}P^j_{03} + 3n_1(P^i_{31}P^j_{02} + P^i_{20}P^j_{13}) + 3n_1(n_1^2 + 2)P^i_{22}P^j_{11} \\
& - 3(2n_1^2 + 1)P^i_{21}P^j_{12} + 3n_3P^i_{11}P^j_{02}P^k_{20} + 6(n_1^3 + 3n_1 - 2)P^i_{11}P^j_{11}P^k_{11}] \\
& - 3n_1SP^i_{11}[S(n_1^2P^i_{22} + P^i_{20}P^j_{02} - n_1^2(P^i_{11})^2 + 2P^i_{11}P^j_{11})] - n_1^3(SP^i_{11})^3]\} \\
& + 3n^{-5}\{S[n_1^2(P^i_{32}B_i + P^i_{23}A_i) + 2\alpha(P^i_{21}P^j_{11}B_i + P^i_{12}P^j_{11}A_i) \\
& - 2n_1(P^i_{21}P^i_{11}B_i + P^i_{12}P^i_{11}A_i) - 2n_1(P^i_{11}P^j_{21}B_i + P^i_{11}P^j_{12}A_i) \\
& + (P^i_{12}P^j_{20}B_i + P^i_{21}P^j_{02}A_i) - 2n_1(P^i_{12}P^j_{20}B_j + P^i_{21}P^j_{02}A_j) \\
& + (P^i_{30}P^j_{02}B_j + P^i_{03}P^j_{20}A_j)]\} + 3n^{-4}\{S[n_1(P^i_{31}B_i^2 + P^i_{13}A_i^2) \\
& + n_2(P^i_{20}P^j_{11}B_i^2 + P^i_{02}P^j_{11}A_i^2) - (P^i_{11}P^i_{20}B_i^2 + P^i_{02}P^i_{11}A_i^2) \\
& - 2(P^i_{20}B_iP^j_{11}B_j + P^i_{02}A_iP^j_{11}A_j) + 2n_1P^i_{22}A_iB_i - 2P^i_{20}B_iP^j_{02}A_j \\
& - 2P^i_{11}A_iP^j_{11}B_j + 2n_2P^i_{11}P^j_{11}A_iB_i - 2(P^i_{11})^2A_iB_i]\} + n^{-3}\{S[(P^i_{30}B_i^3 + P^i_{03}A_i^3) \\
& + 3(P^i_{21}A_iB_i^2 + P^i_{12}A_i^2B_i)]\}. \qquad (14)^{13}
\end{aligned}$$

Where $\alpha = n_1^2 + n_1 + 1$.

This formula is shorter and simpler than the formula for ${}_2M_{p_{21}}$, although they are of the same order. This is due to the symmetry of ${}_3M_{p_{11}}$.

Chapter V. Product Moments of Trivariate and Quadrivariate Populations

1. Some additional definitions and notation. In this chapter we shall indicate briefly how the method of the previous chapters may be extended to populations

[13] Cf. *Biometrika*, Vol. XXI, p. 253, formula (19).

of more than two variables. We shall do this by deriving some of the simpler formulae, corresponding to those of Chapter II, for trivariate and quadrivariate populations.

The notation will be slightly changed in that we shall symbolize the new variables by priming the symbols for the variables used in the previous chapters. Thus, we shall indicate the k^{th} trivariate population by (X_k, Y_k, X'_k) and the k^{th} quadrivariate population by (X_k, Y_k, X'_k, Y'_k), and samples from such populations by (x_k, y_k, x'_k) and (x_k, y_k, x'_k, y'_k) respectively.

We shall denote by P^m_{ijk} the product moment of the m^{th} population of order i in X, j in Y, and k in X', and by P^m_{ijkl} the similar product moment for a quadrivariate population. These are defined by the following equations:

$$P^m_{ijk} = E(X_m - a_m)^i(Y_m - b_m)^j(X'_m - c_m)^k, \tag{5.11}$$

$$P^m_{ijkl} = E(X_m - a_m)^i(Y_m - b_m)^j(X'_m - c_m)^k(Y_m - d_m)^l \tag{5.12}$$

where a_m, b_m, etc. are defined as in Chapter I part 2.

The sample product moments corresponding to P^m_{ijk}, P^m_{ijkl} will be denoted by p_{ijk} and p_{ijkl} respectively. They are defined by:

$$p_{ijk} = n^{-1} \sum_{m=1}^{n} (x_m - x)^i(y_m - y)^j(x'_m - x')^k, \tag{5.13}$$

$$p_{ijkl} = n^{-1} \sum_{m=1}^{n} (x_m - x)^i(y_m - y)^j(x'_m - x')^k(y'_m - y')^l. \tag{5.14}$$

Finally we shall designate $E(p_{ijk})$ and $E(p_{ijkl})$ by $\bar{p}_{ijk}$ and $\bar{p}_{ijkl}$ respectively.

2. The Mathematical Expectation of p_{111} and p_{211}. By definition we have

$$\bar{p}_{111} = E[n^{-1}S(x_i - x)(y_i - y)(x'_i - x')]. \tag{5.21}$$

Applying the transformations (1.17) this equation becomes

$$\begin{aligned} np_{111} &= E[S(U_i + A_i)(V_i + B_i)(U'_i + C_i)] = SE(U_iV_iU'_i) + SE(U_iV_iC_i) \\ &\quad + SE(U_iU'_iB_i) + SE(V_iU'_iA_i) + \text{vanishing terms} + SE(A_iB_iC_i). \end{aligned} \tag{5.22}$$

Since $EA_iB_iC_i = A_iB_iC_i$, $SE(A_iB_iC_i) = SA_iB_iC_i$. Following the previous notation we shall put $SA_iB_iC_i = Q_{111}$.

When the expression $SE(U_iV_iU'_i)$ is expanded, no other non-vanishing terms except those of the form $E(u_iv_iu'_i) = P^i_{111}$ can appear. The coefficient of this term will evidently be the same as that of P^i_{21} in (2.23), namely: $n^{-2}n_1n_2$. Whence:

$$SE(U_iV_iU'_i) = n^{-2}n_1n_2SP^i_{11}.$$

The three terms following the first of (5.22) are by (2.24) equal to

$$n^{-1}n_2S(P^i_{110}C_i + P^i_{101}B_i + P^i_{011}A_i).$$

We thus get:

$$\bar{p}_{111} = n^{-3}n_1n_2SP^i_{111} + n^{-2}n_2S(P^i_{110}C_i + P^i_{101}B_i + P^i_{011}A_i) + n^{-1}Q_{111}. \quad (15)$$

With the aid of the formulae of II, 3 we easily find the formula

$$\begin{aligned}\bar{p}_{212} = n^{-5}\{(n_1^4 - 1)SP^i_{211} + (2n_1^2 + n_2)S(P^i_{200}SP^i_{011} + 2P^i_{110}SP^i_{101} - P^i_{200}P^i_{011}\\ - 2P^i_{110}P^i_{101})\} + n^{-4}\{(n_1^3 + 1)S(P^i_{201}B_i + P^i_{210}C_i + 2P^i_{111}A_i)\}\\ + n^{-3}\{(n_1^2 - 1)S(P^i_{011}A_i^2 + 2P^i_{101}A_iB_i + 2P^i_{110}A_iC_i + P^i_{200}B_iC_i) + Q_{200}SP^i_{011}\\ + 2Q_{110}SP^i_{101} + 2Q_{101}SP^i_{110} + Q_{011}SP^i_{200})\} + n^{-1}Q_{211}. \quad (16)\end{aligned}$$

3. **The Mathematical Expectation of** p_{1111}. The procedure for finding the formula for p_{1111} is very similar to the above. We shall therefore merely state the result.

$$\begin{aligned}\bar{p}_{1111} = n^{-5}\{(n_1^4 - 1)SP^i_{1111} + (2n_1^2 + n_2)S(P^i_{1100}P^j_{0011} + P^i_{1001}P^j_{0110}\\ + P^i_{1010}P^j_{0101})\} + n^{-4}\{(n_1^3 + 1)S(P^i_{1110}D_i + P^i_{1101}C_i + P^i_{1011}B_i + P^i_{0111}A_i)\}\\ + n^{-3}\{(n_1^2 + 1)S(P^i_{1100}C_iD_i + P^i_{1010}B_iD_i + P^i_{0110}A_iD_i + P^i_{0101}A_iC_i\\ + P^i_{0011}A_iB_i + P^i_{1001}B_iC_i + S(Q_{0011}P^i_{1100} + Q_{0101}P^i_{1010} + Q_{1001}P^i_{0110} + Q_{1010}P^i_{0101}\\ + Q_{1100}P^i_{0011} + Q_{0110}P^i_{1001})\} + n^{-1}Q_{1111}. \quad (17)\end{aligned}$$

WASHINGTON UNIVERSITY, ST. LOUIS.

AN APPLICATION OF ORTHOGONALIZATION PROCESS TO THE THEORY OF LEAST SQUARES

By Y. K. Wong

Introduction

The present paper is an outgrowth of the writer's attempt to fill a lacuna in the discussion of the Gauss method of substitution as given by many writers. For illustration, let us cite Brunt's *Combination of Observations*. In Chapter VI, we find:

Let the normal equations be

$$\begin{aligned} [aa]x + [ab]y + [ac]z - [al] &= 0 \\ [bb]y + [bc]z - [bl] &= 0 \\ [cc]z - [cl] &= 0\,. \end{aligned} \tag{i}$$

From this equation we find

$$x = -\frac{[ab]}{[aa]}y - \frac{[ac]}{[aa]}z + \frac{[al]}{[aa]}\,. \tag{ii}$$

Substituting, we obtain

$$\begin{aligned} [bb1]y + [bc1]z - [bl1] &= 0 \\ [cc1]z - [cl1] &= 0 \end{aligned} \tag{iii}$$

where

$$[bb1] = [bb] - [ab]\,[ab]/[aa], \text{ etc.} \tag{iv}$$

From the first equation in (iii),

$$y = -\frac{[bc1]}{[bb1]}z + \frac{[bl1]}{[bb1]}\,. \tag{v}$$

In connection with equations (ii) and (v), the question naturally arises as to whether or not these numbers $[aa]$, $[bb1]$, $\cdots$ are all different from zero. Since $[aa] = \Sigma a_i a_i$, one can see that $[aa] \neq 0$ if $a_i \neq 0$ for every i. However, to show the non-vanishing of $[bb.1]$, $[cc.2]$, etc. is by no means simple. Many writers do not give a demonstration on this point. We know that a system of non-homogeneous linear equations has a solution if the system of equations is linearly independent. Brunt gives a discussion of the independence of the normal equations in Chapter V, Art. 36, but he does not state clearly a condition for independence. He says: "The condition of independence is in general satisfied in

the problems which arise in practice. We can then proceed to the formation and solution of the normal equations." It is one of the aims of this paper to give a necessary and sufficient condition for the independence of the normal equations and to show $[aa]$, $[bb.1]$, etc. are all different from zero when the condition is satisfied.

In the theory of least squares, there is the classical method of the derivation of normal equations by an application of the notion of minimum in differential calculus. After the normal equations are secured, the Gauss method of substitution is applied to obtain the solution. Doolittle modifies the Gauss method of substitution so as to facilitate the labor of computation. However, when the number of parameters (or unknowns) exceeds 4, Doolittle's method is quite complicated. In the present paper the writer wishes to present a mathematical discussion of a method obtained through an application of the Gram-Schmidt orthogonalization process. This method furnishes us a new procedure for determining the most probable values of the parameters (or unknowns). The formulation of the system of normal equations will be omitted in this new procedure, which is particularly effective in fitting curves to time series. The paper can be roughly divided into three parts. The first part gives an algebraic derivation of the normal equations. The second part derives a condition for a set of observation data so that the Gauss method of substitution is applicable. The third part gives a relationship between the Gauss method of substitution and the orthogonalization process. A practical application of the results of this paper will be found in a later paper.

The process of orthogonalization has been used in the 19th century, and has been applied extensively in the theory of integral equations and linear transformations in Hilbert space. In classical analysis, if $\varphi_1(x)$, $\varphi_2(x)$, $\cdots$, defined on $(0, 1)$, is a normally orthogonalized system, and if $f(x)$, defined on $(0, 1)$, is such that f^2 is Lebesgue integrable, then the system of Fourier coefficients

$$f_r = \int_0^1 f(x)\varphi_r(x)dx \qquad (r = 1, 2, \cdots)$$

has certain interesting properties, one of which is that

$$\frac{1}{m}\int_0^1 (f(x) - \sum_1^m f_r\varphi_r)^2\, dx = 0\,.$$

The preceding notion has a close connection with the theory of least squares as outlined in many texts on statistics. In section III, the reader will find how this notion is applied in the derivation of the normal equations. Since the number of dimensions is finite, the integration process reduces to a summation process and furthermore no limiting process is used. This new derivation of normal equations has the advantage that (1) differential calculus is not used, (2) a new form of normal equations is obtained, (3) the solution of the unknowns or parameters can be immediately obtained without further application of the

Gauss Method of Substitution or the Doolittle Method, and (4) the formula for the "quadratic residual" is obtained as a simple corollary.

From the results in section III, we see immediately what condition should be imposed upon the set of observation data so that the Gauss method of substitution may be applicable. In section VI, we find a necessary and sufficient condition for the independence of the system of normal equations (3.9), and also the fact that when this condition is fulfilled, then, due to the special nature of the coefficients of the unknowns, we see that the matrix is properly positive. It is on account of this fact that we are able to show that the numbers $[aa]$, $[bb.1]$, etc. are all different from zero. The demonstration of this point is found in section VII. In this section, we lay down a fundamental hypothesis for Gauss's method of substitution, namely, the set of observations $A_i = (a_{i1}, \cdots, a_{in})$ $i = 1, 2, \cdots, r$, is linearly independent. Lemma 7.3 may be called the fundamental lemma for Gauss's method of substitution. Some interesting properties of the numbers $[A_sA_t \cdot h]$, where $s, t = 1, \cdots, r$, and h is less than the smaller one of (s, t), are demonstrated.

From the properties of the numbers $[A_sA_t \cdot h]$, where $s, t = 1, \cdots, r$ and h is less than the smaller one of (s, t), and in comparison of the system of equations (3.7^0) with the final form of equations obtained through the application of the Gauss method of substitution, we can see the relationship between the Gauss method and the Gram-Schmidt orthogonalization process. If we should like to give some credit to Gauss, we may say that the orthogonalization process was known by him, but was stated in a different form.

The writer wishes to remark that certain theorems together with proofs in section II, IV, V and VI are obtained from E. H. Moore's lecture notes. However the writer should be responsible for any defect. Finally, I should emphasize that the use of the notion of positive matrices is only for convenience.

I. Vectors, Inner Products, and Linear Independence

In this paper, we shall consider vectors of the form[1]

$$(v_1, v_2, \cdots, v_n). \tag{1.10}$$

For convenience, we shall use capital letters to denote vectors of the type (1.10).

Let $V = (v_1, v_2, \cdots, v_n)$ and $U = (u_1, u_2, \cdots, u_n)$, then we say $V = U$ if $v_i = u_i$ for every i.

We define $V + U$ by

$$V + U = (v_1 + u_1, v_2 + u_2, \cdots, v_n + u_n), \tag{1.11}$$

and sV, where s is a number, by

$$sV = (sv_1, sv_2, \cdots, sv_n). \tag{1.12}$$

[1] If we write v_i as $v(i)$, where $i = 1, 2, \cdots, n$, then $v(i)$ may be considered as a function of one variable whose range consists of a set of positive integers, $(1, 2, \cdots, n)$. E. H. Moore defines a vector as a function of one variable.

Hence, $sV = Vs$. In particular, when $s = -1$. we shall put $-V = (-1)V$. Then $U - V$ becomes a special instance of (1.11) and (1.12).

From (1.11) and (1.12), we see that addition is commutative and associative.

INNER PRODUCTS: The inner product of two vectors $V = (v_1, \cdots, v_n)$ and $U = (u_1, \cdots, u_n)$ is defined[2] to be

$$(V, U) = \sum_1^n v_i u_i . \tag{1.2}$$

The norm of a vector V is defined by $n(V) = (V, V)$; and the modulus of a vector V is defined by mod $(V) = +\sqrt{n(V)}$.

From (1.11), (1.12), and (1.2), we can easily prove the following theorem:

THEOREM 1. *The symbol (,) has the following properties:*

(S) $(U, V) = (V, U)$ *for every* V, U; (symmetric property)

(L_s) $(sV, U) = s(V, U) = (V, sU)$ *for every* V, U *and every number* s;

(L_+) $(U, (V + W)) = (U, V) + (U, W)$ *for every* U, V, W; (*linear property*)

(P) $(V, V) \geqq 0$ *for every* V; (*positive property*)

(P_0) $(V, V) = 0$ *if and only if* V *is a zero vector*; (*properly positive property*)

LINEAR INDEPENDENCE. A set of vectors $V_1, \cdots, V_r$ is said to be linearly dependent in case there exist constants $c_1, \cdots, c_r$ not all equal to 0 such that

$$c_1V_1 + \cdots + c_rV_r = 0 ,$$

where 0 is a zero vector.

A set of vectors $V_1, \cdots, V_r$ is said to be linearly independent in case, if the constants $c_1, \cdots, c_r$ satisfy

$$c_1V_1 + \cdots + c_rV_r = 0 ,$$

each constant $c_i = 0$.

THEOREM 2. *If the set* $V_1, \cdots, V_r$ *is linearly independent, then none of the vectors is a zero vector, and hence the norm of every vector must be different from zero.*

For if V_s is a zero vector, then set $c_s = 1$, and $c_i = 0$ for $i \neq s$. It is obvious that

$$0 \cdot V_1 + \cdots + 0 \cdot V_{s-1} + 1 \cdot V_s + 0 \cdot V_{s+1} + \cdots + 0 \cdot V_r = 0 ,$$

which show that the set of vectors $V_1, \cdots, V_r$ is linearly dependent, contradictory to the hypothesis.

A more general theorem is stated in

THEOREM 3. *If the set* $V_1, \cdots, V_r$ *is linearly independent, then every subset*[3] *is also linearly independent.*

We shall prove this theorem by a contrapositive form. The contrapositive form is as follows: *If in the set* $V_1, \cdots, V_r$, *there exists a subset which is linearly*

[2] The notation (,) was introduced by D Hilbert. In treatises on least squares, the notation [ ] is used. The present writer reserves the latter notation for other purposes.

[3] Consider a set of integers $(1, 2, \cdots, n)$ Then any *combination* of this set of n distinct integers taken $r \leqq n$ at a time is called a *subset* of the set $(1, 2, \cdots, n)$. Likewise, we call any combination of the set of vectors $V_1, V_2, \cdots, V_n$ taken $r \leqq n$ at a time a *subset* of the whole set.

dependent, then the whole set is also linearly dependent. Without losing any generality, let us suppose the subset $V_1, \cdots, V_s$ ($s \leqq r$) to be linearly dependent. Then there exist $c_1, \cdots, c_s$ such that

$$c_1V_1 + \cdots + c_sV_s = 0 .$$

If $s = r$, then the whole set is linearly dependent. If $s < r$, then let $c_i = 0$ for $i = s - 1, s - 2, \cdots, r$. Then

$$\sum_1^r c_iV_i = 0 ,$$

which shows the whole set is linearly dependent.

THEOREM 4.[4] *A necessary and sufficient condition for the set $V_i = (v_{i1}, \cdots, v_{in})$, $i = 1, \cdots, r$ to be linearly independent is that there exists a non-vanishing determinant of order r in the array*

$$\begin{array}{l} v_{11}, v_{12}, \cdots, v_{1n} \\ v_{21}, v_{22}, \cdots, v_{2n} \\ \\ v_{r1}, v_{r2}, \cdots, v_{rn} \end{array}$$

II. Gram-Schmidt's Orthogonalization Process

For the present section and the sequel, we shall adopt the notation $A_i = (a_{i1}, \cdots, a_{in})$, $B_i = (b_{i1}, \cdots, b_{in})$, and $C_i = (c_{i1}, \cdots, c_{in})$ for $i = 1, 2, \cdots, r$.

THEOREM 5. *For every set of vectors $A_1, \cdots, A_r$, there exists uniquely a set of vectors $B_1, \cdots, B_r$ such that*

5.1) $(B_t, B_s) = 0$ $(t \neq s)$.

5.2) *For every t satisfying the relation $1 \leqq t \leqq r$, then A_t is a linear combination of $B_1, \cdots, B_t$; and B_t is a linear combination of $A_1, \cdots, A_t$.*

5.3) $B_1 = A_1$; *and for $t > 1$, $(B_t - A_t)$ is a linear combination of $B_1, \cdots, B_{t-1}$, and is also a linear combination of $A_1, \cdots, A_{t-1}$.*

5.4) *If $t > 1$, then $(A_s, B_t) = 0$ for every $s < t$.*

5.5) $(A_t, B_t) = (B_t, B_t) = (B_t, A_t)$ *for every t.*

To prove this theorem, let us define

$$\begin{aligned} B_1 &= A_1 , \\ B_2 &= A_2 \qquad \text{if} \quad n(B_1) = 0 \\ &= A_2 - \frac{(A_2, B_1)}{n(B_1)} B_1 \qquad \text{if} \quad n(B_1) \neq 0 \end{aligned} \tag{2.1}$$

$$B_t = A_t - \sum_{i=1}^{\cdot} h_{ti}B_i \qquad (1 \leqq t \leqq r) ,$$

[4] See Dickson, *Modern Algebraic Theories*, p. 55; Bocher, *Higher Algebra*, p. 36.

where

$$(2.11)\qquad \begin{aligned} h_{t\iota} &= (A_t, B_\iota)/n(B_\iota), && \text{if} \quad n(B_\iota) \neq 0, \\ &= 0, && \text{if} \quad n(B_\iota) = 0. \end{aligned}$$

We proceed to show that this set has the properties stated in the theorem.

To prove 5.1), let us suppose $t < s$. This assumption is permissible since the operator (,) has the symmetric property. First, if $A_1 = 0$, then $B_1 = 0$, and

$$(B_1, B_2) = (A_1, A_2) = (0, A_2) = 0.$$

Secondly, if $A_1 \neq 0$, then $B_1 \neq 0$ and

$$\begin{aligned} (B_1, B_2) &= (A_1, A_2 - h_2, B_1) = (A_1, A_2) - (A_1, B_1)\frac{(A_2, B_1)}{n(B_1)} \\ &= (A_1, A_2) - (A_1, A_1)(A_2, A_1)/n(A_1) = 0. \end{aligned}$$

Assume 5.1) is true for $t = s - 1$, then

$$(B_t, B_s) = \left(B_t, A_s - \sum_1^{s-1} h_{s\iota}B_\iota\right) = (B_t, A_s) - \sum_1^{s-1} h_{s\iota}(B_t, B_\iota).$$

The sum on the right hand side reduces to $h_{st}(B_t, B_t)$, since the other terms vanish by assumption. Now if $(B_t, B_t) \neq 0$ then by (2.11), $h_{st}(B_t, B_t) = (A_s, B_t)$, and by the symmetric property of (,), we obtain

$$(B_t, B_s) = (B_t, A_s) - (A_s, B_t) = 0.$$

If $(B_t, B_t) = 0$, then by the P_0-property of (,), we find that B_t is a zero vector, and hence $(B_t, B_s) = 0$.

5.2) follows from the definition of B_t.

That $(A_t - B_t)$ is a linear combination of $B_1, \cdots, B_{t-1}$ for $t > 1$ follows from the definition of B_t. Since B_s is a linear combination of $(A_1, \cdots, A_{s-1})$, we secure the second part of 5.3).

By 5.2), we can determine $g_{s\iota}$ such that $A_s = \sum_1^s g_{s\iota}B_\iota$. Thus for every $s < t$, we have by 5.1)

$$(A_s, B_t) = \left(\sum_1^s g_{s\iota}B_\iota, B_t\right) = \sum_1^s g_{s\iota}(B_\iota, B_t) = .0$$

By 5.3), there exist $g_{t\iota}$ such that $A_t - B_t = \sum_1^{t-1} g_{t\iota}B_\iota$ and hence $A_t = B_t + \sum_1^{t-1} g_{t\iota}B_\iota$. Thus by 5.1), we have

$$\begin{aligned} (A_t, B_t) &= \left(B_t + \sum_1^{t-1} g_{t\iota}B_\iota, B_t\right) = (B_t, B_t) + \sum_1^{t-1} g_{t\iota}(B_\iota, B_t) \\ &= (B_t, B_t). \end{aligned}$$

By the symmetric property of (,), we secure $(A_t, B_t) = (B_t, B_t)$.

For the proof of uniqueness, let us suppose there exists a second set of vectors $B'_1, \cdots, B'_r$ having the properties 5.1), 5.2), 5.3), 5.4), and 5.5). By 5.3), we see that $B_1 = A_1 = B'_1$. Assuming the uniqueness holds true for $r = t$, we proceed to show that it is also true for $r = t + 1$. By 5.3) there exist constants s_i, s'_i $(i = 1, \cdots, t)$ such that

$$B_{t+1} = A_{t+1} + \sum^{t} s_i A_i$$

$$B'_{t+1} = A_{t+1} + \sum_{1}^{t} s'_i A_i .$$

Thus

$$B_{t+1} - B'_{t+1} = \sum_{1}^{t} (s_i - s'_i) A_i .$$

From this, we secure

$$(B_{t+1} - B'_{t+1}, B_{t+1} - B'_{t+1}) = \left(B_{t+1} - B'_{t+1}, \sum_{1}^{t} (s_i - s'_i) A_i\right)$$

$$= \sum_{1}^{t} (s_i - s'_i) \cdot (B_{t+1} - B'_{t+1}, A_i) = 0 ,$$

by virtue of 5.4). Hence by P_0-property of (,), we have $B_{t+1} - B'_{t+1} = 0$ and hence $B_{t+1} = B'_{t+1}$.

The set $B_1, \cdots, B_r$ with the properties stated in Theorem 5 is called the *orthogonalized set* of $A_1, \cdots, A_r$. This process is called Gram-Schmidt's orthogonalization process.

The set $B_1, \cdots, B_r$ is called the *normally orthogonalized set* of $A_1, \cdots, A_r$ in case the former set enjoys the properties 5.1), 5.2), 5.3), 5.4), and if

5.5n) $\qquad (A_t, B_t) = (B_t, B_t) = (B_t, A_t) = 1$ *for every* t .

THEOREM 6. *If a subset $A_{k_1}, \cdots, A_{k_m} (1 \leqq k_1 \leqq \cdots \leqq k_m \leqq r)$ in the set $A_1, \cdots, A_r$, is linearly independent, then there is a subset $B_{k_1}, \cdots, B_{k_m}$ which has the properties stated in Theorem 5, and it is also linearly independent.*

Let $h = k_m - k_1 + 1$. To prove the theorem, we may assume $k_1, \cdots, k_m$ to be $1, \cdots, h \leqq r$, for otherwise, we may renumber the vectors. We construct the B vectors in the same way as given in equation (2.1) and (2.11). By Theorem 5, we have

$$B_1 = A_1, \qquad B_s = A_s + \sum_{1}^{s-1} g_{si} A_i \qquad (s = 2, \cdots, h) . \tag{2.2}$$

Suppose the constants $c_1, \cdots, c_h$ be such that

$$c_1 B_1 + \cdots + c_h B_h = 0 .$$

Then by (2.2), we secure

$$0 = c_1A_1 + \sum_2^h c_sB_s = c_1A_1 + \sum_2^h c_s\left(A_s + \sum_1^{s-1} g_{si}A_i\right)$$

$$= (c_1 + c_2g_{21}, + \cdots + c_hg_{h1})\,A_1 + (c_2 + c_3g_{32} + \cdots c_hg_{h2})A_2 + \qquad + c_hA_h\,.$$

Since $A_1, \cdots, A_h$ are linearly independent, we have

$$(2.3)\qquad \begin{aligned} c_1 - c_2g_{21} - \cdots - c_hg_{h1} &= 0\,, \\ c_2 - \cdots - c_hg_{h2} &= 0\,, \\ \\ c_h &= 0\,. \end{aligned}$$

But the determinant of the coefficients of $c_i(i = 1, \cdots, h)$ is

$$\begin{matrix} 1 & g_{21} & g_{31} & \cdots & g_{h1} \\ 0 & 1 & g_{32} & \cdots & g_{h2} \\ \\ 0 & 0 & 0 & & 1 \end{matrix}$$

Hence by a theorem in the theory of equations,[5] the only solution that satisfies (2.3) is that $k_1 = k_2 = \cdots = k_h = 0$. Thus the subset $B_1, \cdots, B_h$ is linearly independent.

COROLLARY. *The orthogonalized set $B_1, \cdots, B_r$ is linearly independent if and only if the set $A_1, \cdots, A_r$ is linearly independent.*

THEOREM 7a. *If a set of vectors $A_1, \cdots, A_r$ is linearly independent, then the set can be normally orthogonalized.*

Let B_i be the orthogonalized set of A_i. Since A_i is a linearly independent set, then the set B_i is also linearly independent by Theorem 6. Hence by Theorem 2, the norm of every vector B_i is non-vanishing. Define $C_i = B_i/\text{mod}\,(B_i)$. Then this set C_i enjoys the properties 5.1), 5.2), 5.3), 5.4) and 5.5n).

THEOREM 7b. *If a set of vectors, $V_1, \cdots, V_r$ is normally orthogonal, i.e.* if

$$(2.4)\qquad (V_i, V_j) = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j)\,, \end{cases}$$

then $V_1, \cdots, V_r$ is linearly independent.

For suppose

$$c_1V_1 + \cdots + c_rV_r = 0\,.$$

Then

$$\sum^r c_i(V_i, V_j) = 0\,, \qquad (j = 1, 2, \cdots, r)\,.$$

[5] Dickson, *First Course in the Theory of Equations* (1922), p. 119.

By condition (2.4), the preceding expression reduces to

$$c_j = 0, \qquad (j = 1, 2, \cdots, r),$$

which shows the linear independence of $V_1, \cdots, V_r$.

III. Algebraic Derivation of the Normal Equations

Consider a linear function

$$l = p_1x_1 + p_2x_2 + \cdots + p_rx_r = \sum_1^r p_ix_i. \tag{3.1}$$

Let the set of observations of x_i and l be

$$A_i = (a_{i1}, \cdots, a_{in}), \qquad L = (l_1, \cdots, l_n) \qquad (i = 1, \cdots, r; n \geqq r) \tag{3.2}$$

respectively, then the residual v_i is

$$v_i = \sum_{j=1}^r p_ja_{ji} - l_i, \qquad (i, = 1, \cdots, n).$$

In vector notation,

$$V = \sum_{j=1}^r p_jA_j - L.$$

The theory of least squares requires us to find the values for $p_1, \cdots, p_r$ so as to make (V, V) a minimum, or

$$(\sum p_jA_j - L, \sum p_jA_j - L) = \text{a minimum}. \tag{3.3°}$$

Let $A_1, \cdots, A_r$ be linearly independent. By Theorem 7, the vectors $A_1, \cdots, A_r$ can be normally orthogonalized. Let $C_1, \cdots, C_r$ be the normally orthogonal set. Then every A_t $(t = 1, \cdots, r)$ is expressible as a linear combination of $C_1, \cdots, C_t$. Let us write

$$\sum_1^r p_jA_j = \sum_1^r k_jC_j. \tag{3 3}$$

Our problem now is equivalent to that of finding the values $k_i(i = 1, \cdots, r)$ so as to render the inner product

$$(\sum k_jC_j - L, \sum k_jC_j - L) \tag{3.4}$$

a minimum. Expression (3.4) can be written in the form

$$\begin{aligned} &(L, L) - 2\sum(L, C_i)k_i + \sum_{i,j}(k_iC_i, k_jC_j) \\ &= (L, L) - 2\sum(L, C_i)k_i + \sum k_i^2 \\ &= (L, L) - \sum(L, C_i)^2 + \sum(k_i - (C_i, L))^2. \end{aligned} \tag{3.5}$$

Hence (3.4) gives a minimum if and only if the last summation vanishes, i.e.,

$$k_i = (C_i, L) \qquad (i = 1, \cdots, r). \tag{3.6}$$

The Bessel's inequality

$$\sum_{1}^{r} k_i^2 \leqq (L, L)$$

is obtained from (3.6), (3.4), and (3.5).

To solve for p_i, we make use of (3.3) and (3.6), and secure

$$\sum_{1}^{r} A_i p_i = \sum_{1}^{r} (C_i, L) C_i ,$$

whence

$$\left(C_k, \sum_{1}^{r} A_j p_j\right) = \left(C_k, \sum_{1}^{r} (C_i, L)\, C_i\right).$$

On the right hand side we have

$$(C_k, \sum (C_i, L) C_i) = \sum (C_i, L)\, (C_k, C_i) = (C_k, L) ,$$

since $(C_k, C_i) = 0$ when $i \neq k$, and $(C_k, C_i) = 1$ when $i = k$. On the left hand side, we have

$$\left(C_k, \sum_{j=1}^{r} A_j p_j\right) = \sum_{j=1}^{r} (C_k, A_j) p_j = \sum_{j=k}^{r} (C_k, A_j) p_j ,$$

since $(C_k, A_j) = 0$ when $j < k$. Hence the values for $p_1, \cdots, p_r$ are given by

$$\sum_{i=k}^{r} (C_k, A_i) p_i = (C_k, L) \qquad (k = 1, \cdots, r) , \tag{3.7}$$

where $(C_i, A_i) = (C_i, C_i) = 1$.

Equations (3.7) are called the normal equations, which are derived without using any notion in differential calculus.

From (3.6) and (3.5), we secure the value for the 'quadratic residual' (V, V):

$$(V, V) = (L, L) - \sum_{i=1}^{r} (L, C_i)^2, \tag{3.8}$$

which is a positive quantity by virtue of the Bessel's inequality.

Let $B_1, \cdots, B_r$ be an orthogonalized set of $A_1, \cdots, A_r$. Then every vector B_i has a non-vanishing norm, and $B_i = \operatorname{mod} (B_i) \cdot C_i$. Hence from (3.7) and (3.8), we have

$$\sum_{i=k}^{r} (B_k, A_i) p_i = (B_k, L), \qquad (k = 1, 2, \cdots, r) , \tag{3.7°}$$

$$(V, V) = (L, L) - \sum_{i=1}^{r} (L, B_i)^2 / n(B_i) . \tag{3.8°}$$

Thus we have proved the following

THEOREM 8. *Given a linear function* (3.1). *Let the set of observations of* x_i *and* l *be*

$$A_i = (a_{i1}, \cdots, a_{in}), \qquad L = (l_1, \cdots, l_n) \qquad (i = 1, \cdots r;\, n \geqq r)$$

respectively. Let $A_1, \cdots, A_r$ be linearly independent, $B_1, \cdots, B_r$ be the orthogonalized set, and $C_1, \cdots, C_r$, the normally orthogonalized set of $A_1, \cdots, A_r$. Then the set of values $p_1, \cdots, p_r$ will minimize (3.3°) if and only if the system of equations (3.7°) or (3.7) holds true; in other words, $\sum_{i=1}^{r} p_i A_i - L$ is orthogonal to C_j or to B_j for every j. The quadratic residual (V, V) is given by (3.8°) or (3.8).

From (3.7), we can secure the solution for $p_1, \cdots, p_r$ immediately without further application of the Gauss method of substitution.

The proof of the following theorem does not make use of the orthogonalization process.[6]

THEOREM 8°. *Let $F = \Sigma p_i A_i$, where every A_i is not a zero vector. The set of values $p_1, \cdots, p_r$ will minimize (3.3°) if and only if $(F - L, A_i) = 0$ for every i*, i.e., *$F - L$ is orthogonal to every A_i.*

The condition is necessary. To prove this, we show that if $(F - L, A_i) \neq 0$ for every i, then we can find another set $q_1, \cdots, q_r$ such that $n(F - L) > n(G - L)$, where $G = \Sigma q_i A_i$. For if $(F - L, A_i) \neq 0$ for every i, then we can find a vector A_s such that $(F - L, A_s) \neq 0$. Since $A_s \neq 0$, we let $e = (F - L, A_s)/n(A_s)$ and $G = F - eA_s = \Sigma q_i A_i$. Then

$$n(G - L) = n(F - eA_s - L) = n(F - L) - (F - L, A_s)^2/n(A_s)\,,$$

which shows that $n(G - L) < n\,(F - L)$.

To prove the sufficiency, we show that for every set $q_1, \cdots, q_r$ different from $p_1, \cdots, p_r$ then $n(G - L) > n(F - L)$, where $G = \Sigma q_i A_i$. Let $s_i = q_i - p_i$, and $H = \Sigma s_i A_i$. Then $G = F + H$. Now if $(F - L, A_i) = 0$ for every i, it follows that

$$(F - L, H) = \sum_{i=1} (F - L, A_i)s_i = 0\,.$$

Thus

$$n(G - L) = n(F - L) + n(H)\,.$$

Since $n(H) > 0$, we have $n(G - L) > n(F - L)$.

The preceding theorem does not require the linear independence of the vectors $A_1, \cdots, A_r$. By Theorem 7a and 7b we see that it is necessary and sufficient for the set $A_1, \cdots, A_r$ to be linearly independent in order to solve the equations $(F - L, A_i) = 0$, $(i = 1, 2, \cdots, r)$, or

$$\begin{array}{ll} & (A_1, A_1)p_1 + (A_1, A_2)p_2 + \cdots + (A_1, A_r)p_r = (A_1, L) \\ (3.9) & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ & (A_r, A_1)p_1 + (A_r, A_2)p_2 + \cdots + (A_r, A_r)p_r = (A_r, L)\,. \end{array}$$

[6] The proof is based on the same type of reasoning as used by Jackson. See Dunham Jackson's *Theory of Approximation*, pp. 151–152.

If $A_1, \cdots, A_r$ are linearly independent, the conclusion in Theorem 8° can be deduced from Theorem 8. For by Theorem 7a) $A_i = \sum s_{it} C_t$, and hence

$$(F - L, A_i) = \left(F - L, \sum_t s_{it} C_t\right) = \sum_t s_{it}(F - L, C_t) = 0 .$$

Also, Theorem 8 can be deduced from Theorem 8°.

IV. Matrices and Their Reciprocals

An ordered array of numbers of the form

$$\alpha = (a_{ij}) = \begin{array}{lll} a_{11}, & a_{12}, & a_{1m} \\ a_{21}, & a_{22}, & a_{2m} \\ & & \\ a_{n1}, & a_{n2}, & \end{array} \tag{4.1}$$

is a matrix. If we write $a(i, j) = a_{ij}$, then the array of numbers (4.1) may be considered as a function of two variables i, j on the ranges of positive integers $(1, 2, \cdots, n)$, $(1, 2, \cdots, m)$.[7] Thus a vector is a special instance of a matrix. We shall use Greek letters to denote matrices throughout this paper unless otherwise specified. When $n = m$, i.e. the number of rows is the same as the number of columns, we have a square matrix. Associated with every n-row square matrix, κ, a determinant can be defined, and for simplicity, we shall adopt the following notation:

$$D(\kappa) = \begin{array}{ll} a_{11} & a_{1n} \\ & \\ a_{n1} & \end{array}$$

An identity matrix, denoted by $\delta = (d_{ij})$, is a square matrix of which the elements in the principal diagonal are 1 and elsewhere 0, i.e. $d_{ij} = 0$ $(i \neq j)$, $d_{ii} = 1$. A zero matrix, indicated by ω, is one such that every one of its elements is 0. The transposed matrix, α', of α is formed by interchanging the rows and columns. We say two matrices $\alpha = (a_{ij})$ and $\beta = (b_{ij})$ are equal in case $a_{ij} = b_{ij}$ for every i, j. A matrix α is symmetric in case $\alpha' = \alpha$. The i^{th} column of α is indicated by $\alpha(., i)$, the i^{th} row of β by $\beta(i, .)$ and the element in the i^{th} row and j^{th} column by $\alpha\ (i, j)$. Hence $\alpha(i, j) = a_{ij}$.

Addition: Let α be a matrix given by (1) and $\beta = (b_{ij})$ a matrix of the same number of rows and columns as α. Then

$$\alpha + \beta = (a_{ij} + b_{ij}) .$$

[7] E. H. Moore defines a matrix as a function of two variables.

We note that $\alpha + \beta = \beta + \alpha$. If γ is a matrix of the same number of rows and columns as α, then $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

MULTIPLICATION: Let $\alpha = (a_{ij})$ be defined by (1), and $\beta = (b_{jk})$ be a matrix of m row and r columns, then the product $\pi = \alpha\beta$ is defined by

$$\pi = (p_{ik}) = \left(\sum_{j=1}^{m} a_{ij}b_{jk}\right).$$

Thus π is a matrix of m rows and r columns.

The multiplication of two matrices is not necessarily commutative.

If α is a matrix of n rows and m columns, β of m rows and r columns, and γ of r rows and s columns, then $\alpha(\beta\gamma) = (\alpha\beta)\gamma$. If α is a matrix of n rows and m columns, and β, γ are matrices of m rows and r columns, then $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.

SCALAR MULTIPLICATION: Let s be a number, and α be a matrix of n rows and m columns, then

$$s \cdot \alpha = (sa_{ij}) = \alpha \cdot s.$$

Let δ_s denote a square matrix of n rows in which the elements in the principal diagonal are s, and 0 elsewhere. Then $\delta_s = s\delta$, where δ is an n row identity matrix. We note from the associative law of multiplication that

$$s\alpha = \delta_s \cdot \alpha = \alpha \cdot \delta_s.$$

In particular, let $s = -1$, then we have -1α. For convenience, we write $-\alpha = -1\alpha$. From the definition of addition, we obtain a definition of subtraction for two matrices of the same number of rows and columns.

RECIPROCALS OF MATRICES: Let α be a matrix of n rows and m columns. Then a matrix α^{-1} of m rows and n columns is said to be a reciprocal of α in case

$$\alpha \cdot \alpha^{-1} = \delta^n, \quad \text{and} \quad \alpha^{-1} \cdot \alpha = \delta^m,$$

where δ^n, δ^m are identity matrices of order n, m respectively. If a matrix α has a reciprocal α^{-1}, we can prove α^{-1} is unique. It can be shown that *when α has a reciprocal, it must be a square matrix.*[8]

A matrix is said to be non-singular in case it has a reciprocal, otherwise it is said to be singular.[9] It is evident that every zero matrix is singular, and an identity matrix is non-singular.

Suppose α is a square matrix of order n. Let us denote the cofactor of the element a_{ij} of α by e_{ji}. Then

$$\epsilon = (e_{ij}) = \begin{pmatrix} e_{11} & \cdots & e_{1n} \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ e_{n1} & \cdots & e_{nn} \end{pmatrix}$$

is called the adjoint matrix of α.

[8] For the proof of this statement, see Moore, *Vector, Matrices, and Quaternions.*

[9] This definition is due to E. H. Moore.

If α is symmetric, then ϵ is also symmetric. Since $a_{i1}e_{1j} + \cdots + a_{in}e_{nj} = D(\alpha)$ or 0 according as $i = j$ or $i \neq j$, we secure the following:

THEOREM 9. *Let α be a square matrix and ϵ its adjoint, then*

$$\alpha\epsilon = \epsilon\alpha = [D(\alpha)]\delta .$$

THEOREM 10. *If the determinant of α is different from zero, then there exists a reciprocal α^{-1}, and $\alpha^{-1} = adj\ \alpha/D(\alpha)$.*

This theorem follows from theorem 5.

The converse of Theorem 6 is also true.

V. Symmetric Matrices of Positive Type[10]

Let $\alpha = (a_{ij})$ be a matrix of n rows and m columns; and let $\sigma = (k_1, \cdots, k_n)$ and $\rho = (h_1, \cdots, h_m)$ be integers among the sets $(1, \cdots, n)$ and $(1, \cdots, m)$ respectively. The subsets σ and ρ may be equal to the whole sets $(1, \cdots, n)$ and $(1, \cdots, m)$ respectively. Then

$$\alpha(\sigma, \rho) = \begin{matrix} a_{k_1h_1} & \cdots & a_{k_1h_n} \\ \\ a_{k_nh_1} & & a_{k_nh_m} \end{matrix} \tag{3}$$

is called a minor of α. In notation we write this minor as $\alpha(\sigma, \rho)$ indicating the ranges to be σ and ρ.

The minor $\alpha(-\sigma, -\rho)$, which is obtained by striking out all the k_ith $(i = 1, \cdots, m)$ columns and h_jth $(j = 1, \cdots, m)$ rows from α, is called the complementary minor of α (σ, ρ).

If α is a square matrix of order n, then $\alpha(\sigma, \sigma)$ is called a principal minor of α.

Let α and β be matrices of n rows and m columns; and let σ, ρ have the same meaning as above. Then $\alpha(\sigma, \rho)$, $\beta(\sigma, \rho)$ are called corresponding minors in α, β respectively.

A symmetric matrix $\alpha = (a_{ij})$ of order n is said to be of *positive type* in case the determinant of every principal minor of α is positive, and is said to be of *properly positive type* in case the determinant of every principal minor of α is greater than zero.

COROLLARY V1. *Every element in the principal diagonal of a positive, symmetric matrix is positive.*

For, let σ consist of a single integer i, then $\alpha(\sigma, \sigma) = a_{ii} \geqq 0$.

COROLLARY V2. *If a symmetric matrix is properly positive, then every element in the principal diagonal is greater than* 0.

THEOREM 11. *If a symmetric matrix α of order n is (properly) positive, then its adjoint matrix ϵ is also symmetric and (properly) positive.*

[10] We follow the terminology of E. H. Moore. Moore developed this notion quite extensively.

The symmetry of ϵ is evident. Let σ be a subset of $(1, \cdots, n)$ and let p be the number of integers in σ. Consider any principal minor $\epsilon(\sigma, \sigma)$ in the adjoint matrix ϵ. By a theorem in the theory of determinants, we have[11]

$$D[\epsilon(\sigma, \sigma)] = (-1)^{2k} \cdot D[\alpha(-\sigma, -\sigma)] \cdot [D(\alpha)]^{p-1},$$

where k is an integer depending on the set σ. By hypothesis α is positive (properly positive); hence $D[\alpha(-\sigma, -\sigma)]$ and $[D(\alpha)]^{p-1}$ are positive (greater than 0), and it follows that $D[\epsilon(\sigma, \sigma)]$ is positive (greater than 0).

THEOREM 12. *If a symmetric matrix is properly positive, then $D(\alpha)$ is different from zero, and α has a reciprocal α^{-1}, which is also symmetric and properly positive.*

For take σ to be the whole set $(1, \cdots, n)$ in the definition of proper positiveness, and we see that $D(\alpha) \neq 0$. The theorem now follows from Theorems 10 and 11.

VI. Gramian Matrices

In this section, we shall study the matrices of the normal equations (3.9). The main result is that if the set of observations $A_1, \cdots, A_r$ is linearly independent, then the matrix (called Gramian matrix) is properly positive and has a reciprocal which is also properly positive.

THEOREM 13. *Let $A_1, \cdots, A_r$ be a set of vectors, and let $B_1, \cdots, B_r$ be the orthogonalized set of vectors. Then the matrix*

$$\zeta(A_1, \cdots, A_r) = \begin{pmatrix} (A_1, A_1) & \cdots & (A_1, A_r) \\ \cdots & \cdots & \cdots \\ (A_r, A_1) & \cdots & (A_r, A_r) \end{pmatrix} \tag{6.1}$$

has the following properties:

13.1) *symmetry*

13.2) $D[\zeta(A_1, \cdots, A_r)] = n(B_1)n(B_2) \cdots n(B_r)$,

13.3) *positiveness.*

A matrix of the form (6.1) is called a Gramian matrix.

In fact, the symmetric property follows from the fact that $(A_i, A_j) = (A_j, A_i)$ for every i, j.

We shall prove 13.2) by induction. For $r = 1$, we have by Theorem 5

$$(A_1, A_1) = (B_1, B_1) = n(B_1).$$

Assume the equality is true for $r = t$, we shall show it is true for $r = t + 1$. The $(t + 1)$-row determinant is as follows:

$$D[\zeta(A_1, \cdots, A_r)] = \begin{vmatrix} (A_1, A_1) & \cdots & (A_1, A_t) & (A_1, A_{t+1}) \\ \cdots & \cdots & \cdots & \cdots \\ (A_1, A_t) & \cdots & (A_t, A_t) & (A_t, A_{t+1}) \\ (A_1, A_{t+1}) & \cdots & (A_t, A_{t+1}) & (A_{t+1}, A_{t+1}) \end{vmatrix} \tag{6.2}$$

[11] In case $\sigma = (1, \cdots, n)$, $-\sigma$ is a null class Λ (a class which contains no element); then we define $D[\alpha(-\sigma, -\sigma)] = 1$. For the proof of this theorem, see Bocher, p. 31.

By Theorem 5, there exist constants $s_i (i = 1, \cdots, t)$ such that

$$A_{t+1} = B_{t+1} + \sum_{i=1}^{t} s_i A_i .$$

Substituting this value into the last row, we find the element in the i^{th} column is

$$(A_i, A_{t+1}) = \left(A_i, B_{t+1} + \sum_{j=1}^{t} s_j A_j\right) = (A_i, B_{t+1}) + \sum_{j=1}^{t} s_j (A_i, A_j)$$
$$(i = 1, \cdots, t, t+1).$$

The second term on the right is a linear combination of the first t elements in the i^{th} column of the determinant (6.2) and hence by the theory of determinants,[12] we secure

$$D[\zeta(A_1, \cdots, A_{t+1})] = \begin{matrix} (A_1, A_1) & \cdots (A_1, A_t) & (A_1, A_{t+1}) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ (A_1, A_t) & \cdots (A_t, A_t) & (A_t, A_{t+1}) \\ (A_1, B_{t+1}) & \cdots (A_t, B_{t+1}) & (A_{t+1}, B_{t+1}) \end{matrix}$$

By Theorem 5, we find that $(A_i, B_{t+1}) = 0$ for $i = 1, \cdots, t$, and $(A_{t+1}, B_{t+1}) = (B_{t+1}, B_{t+1})$, and hence the preceding determinant reduces to a form in which the first t elements in the $(t+1)^{\text{th}}$ row are zero. Thus

$$D[\zeta(A_1, \cdots, A_{t+1})] = \left|\begin{matrix} (A_1, A_1) & \cdots & (A_1, A_t) \\ (A_1, A_t) & \cdots & (A_t, A_t) \end{matrix}\right| \cdot n(B_{t+1})$$
$$= n(B_1) n(B_2) \cdots n(B_t) n(B_{t+1})$$

which proves 13.2).

Consider any subset $\sigma = (k_1, \cdots, k_m)$ of the set $(1, \cdots, r)$. By the same argument as above, we find that the determinant of any principal minor

$$(6.3) \qquad \begin{matrix} (A_{k_1}, A_{k_2}) & \cdots & (A_{k_1}, A_{k_m}) \\ \cdots\cdots & \cdots & \cdots\cdots \\ (A_{k_m}, A_{k_1}) & \cdots & (A_{k_m}, A_{k_m}) \end{matrix} = n(B_{k_1}) \cdots n(B_{k_m}).$$

By Theorem 1, the number on the right is positive. Thus the matrix ζ is positive.

THEOREM 14. *The following three assertions are equivalent:*

14.1) *the set* $A_1, \cdots, A_r$ *is linearly independent*;

14.2) *the Gramian matrix* (6.1) *is properly positive*;

14.3) *The determinant of the Gramian matrix* (6.1) *is different from zero.*

We shall prove that 14.1) implies 14.2); 14.2) implies 14.3); and 14.3) implies 14.1). We thus prove the three statements are equivalent.

[12] Dickson, *First Course in the Theory of Equations* (1922), p. 113.

Let $B_1, \cdots, B_r$ be the orthogonalized set of the set $A_1, \cdots, A_r$. Since the set $A_1, \cdots, A_r$ is linearly independent, then every subset

$$A_{k_1}, \cdots, A_{k_m} (1 \leqq k_1 \leqq \cdots \leqq k_m \leqq r)$$

is also linearly independent, and hence $n(B_{k_i}) > 0$ for $i = 1, 2, \cdots, m$. By the same argument as given in the demonstration of Theorem 11, we find that the determinant of any principal minor (6.3) is greater than zero. This proves the matrix (6.1) is properly positive.

If the matrix (6.1) is propeıly positive, then by Theorem 10 the determinant of (6.1) is different from zero.

To prove 14.3) implies 14.1), suppose $k_i (i = 1, \cdots, r)$ are such that

$$k_1 A_1 + \cdots + k_r A_r = 0.$$

Then

$$(k_1 A_1 + \cdots + k_r A_r, A_i) = k_1(A_1, A_i) + \cdots + k_r(A_r, A_i) = 0$$

for $i = 1, \cdots, r$. Since $(A_i, A_j) = (A_j, A_i)$, and $D(\zeta) \neq 0$, the set of constants k_i must be all equal to 0.[13]

From Theorem 14, and Theorem 10, we may state the following

COROLLARY: *If the set of observations $A_1, \cdots, A_r$ is linearly independent, then the Gramian matrix ζ has a reciprocal which is properly positive.*

VII. Gauss Method of Substitution

LEMMA 7.1) *Let $\varphi = (s_{ij})$ be an r-row symmetric matrix such that $s_{11} \neq 0$. Then there exists an r-row square matrix τ whose determinant is unity such that $\psi = (r_{ij}) = \tau\varphi$ has the following properties:*

a) *$r_{i1} = 0$ for $i > 1$, and $r_{1i} = s_{1i}$ for every i;*

b) *the first minor of r_{11} is symmetric;*

c) *the determinant of every principal minor in ψ of the form*

$$\begin{matrix} s_{11} & s_{1k_2} & \cdots & s_{1k_m} \\ 0 & r_{k_2k_2} & \cdots & r_{k_2k_m} \\ \\ 0 & r_{k_2k_m} & \cdots & r_{k_mk_m} \end{matrix} \qquad (2 \leqq k_2 \leqq \cdots \leqq k_m \leqq r) \tag{7.1}$$

is equal to the determinant of the corresponding principal minor in φ.

To prove this lemma, let us define

$$\tau = \delta + F_1 \cdot D_1, \tag{7.2}$$

where D_1 is the first row of an r-row identity matrix δ, and $F_1(1) = 0$,

$$F_1(n) = -s_{1n}/s_{11} \qquad (n > 1).$$

(Thus $F_1 D_1$ is an r-row square matrix in which the first column is F_1 and everywhere else 0.) It is clear that τ thus defined is a square matrix of order r, and

[13] See footnote 5.

$D(\tau) = D(\delta + F_1 D_1) = 1$. By multiplication of these two matrices, $\tau\varphi$, we obtain a new matrix such that $r_{11} = s_{11}$, $r_{i1} = 0$ for $i > 1$, and $r_{1i} = s_{1i}$ for every i, and further

$$(7.3) \qquad r_{ij} = s_{ij} - s_{1i}\cdot s_{1j}/s_{11} \text{ for } i > 1,\, j > 1\,.$$

To prove property (b), we note that $s_{ij} = s_{ji}$, since φ is symmetric. Thus for $i > 1, j > 1$, we note from 7.3) that

$$r_{ij} = s_{ij} - s_{1i}s_{1j}/s_{11} = s_{ji} - s_{1j}s_{1i}/s_{11} = r_{ji}\,.$$

For the proof of the last property, we note that the corresponding minor of (7.1) in φ is of the form

$$(7.4) \qquad \begin{bmatrix} s_{11} & s_{1k_2} & \cdot & s_{1k_m} \\ s_{1k_2} & s_{k_2k_2} & \cdot & s_{k_2k_m} \\ \vdots & & & \\ s_{1k_m} & s_{k_2k_m} & \cdots & s_{k_mk_m} \end{bmatrix}$$

Since φ is symmetric, we have by (7.3),

$$r_{k_ik_j} = s_{k_ik_j} - s_{1k_i}s_{1k_j}/s_{11} \qquad (i > 1,\, j > 1),$$

$$0 = s_{k_i1} - s_{1k_i}s_{11}/s_{11} \qquad (i > 1).$$

Thus by a theorem in the theory of determinants, the determinants of (7.1) and (7.4) are equal.

LEMMA 7.2) *Let $\varphi = (s_{ij})$ $(i, j = 1, \cdots, r)$ be a symmetric matrix of positive type, and $s_{11} \neq 0$. Then there exists an r-row square matrix τ whose determinant is unity such that $\psi = (r_{ij}) = \tau\varphi$ has the properties stated in Lemma* 7.1) *and furthermore the minor of r_{11} in* 7.1) *is of positive type.*

To prove the positiveness of the minor of r_{11}, let the determinant of any one of its principal minors be

$$M_1 = \begin{vmatrix} r_{k_2k_2} & \cdots & r_{k_2k_m} \\ \vdots & & \vdots \\ r_{k_2k_m} & \cdots & r_{k_mk_m} \end{vmatrix} \qquad (2 \leqq k_1 \leqq \cdots \leqq k_m \leqq r),$$

where $r_{k_ik_j} = r_{k_jk_i}$ $(i, j = 2, \cdots, m)$ due to the symmetry. Now consider the bordered determinant

$$M_2 = \begin{vmatrix} r_{11} & r_{1k_2} & & r_{1k_m} \\ 0 & r_{k_2k_2} & \cdots & r_{k_2k_m} \\ & \vdots & & \\ & r_{k_2k_m} & & r_{k_mk_m} \end{vmatrix}$$

which by property (a) in Lemma 7.1) gives $M_2 = r_{11}M_1 = s_{11}M_1$. By property (c) in Lemma 7.1), M_2 is equal to the determinant of the form (7.4), which by hypothesis is positive. Thus $s_{11}M_1 \geqq 0$. Since $s_{11} > 0$, we conclude that $M_1 = M_2/s_{11} \geqq 0$.

Lemma 7.3). *Let $\varphi = (s_{ij})$ $(i, j = 1, 2, \cdots, r)$ be a symmetric matrix of properly positive type. Then there exists an r-row square matrix τ whose determinant is unity such that $\psi = (r_{ij}) = \tau\varphi$ has the properties stated in Lemma 7.1) and furthermore the minor of r_{11} in ψ is properly positive.*

Since φ is properly positive, we find that $s_{11} > 0$. The proof of this lemma is similar to that of Lemma 7.2).

Suppose that the set of observations $A_1, \cdots, A_r$ is linearly independent. Then by Theorem 14, the Gramian matrix (6.1) is symmetric and properly positive, and hence $(A_1, A_1) > 0$. By Lemma 7.3), the matrix (6.1) can be reduced to the form

$$\text{(7.5)} \qquad \begin{bmatrix} [A_1A_1\cdot 0] & [A_1A_2\cdot 0] & \cdots\cdots\cdots\cdots & [A_1A_r\cdot 0] \\ 0 & [A_2A_2\cdot 1] & [A_2A_3\cdot 1] \ \cdots & [A_2A_r\cdot 1] \\ 0 & [A_2A_r\cdot 1] & [A_3A_r\cdot 1] \ \cdots & [A_rA_r\cdot 1] \end{bmatrix}$$

where

$$[A_1A_t\cdot 0] = (A_1, A_t) = [A_tA_1\cdot 0] \qquad (t = 1,$$

$$[A_tA_s\cdot 1] = \frac{[A_1A_1\cdot 0]\,[A_tA_s\cdot 0] - [A_1A_t\cdot 0]\,[A_1A_s\cdot 0]}{[A_1A_1\cdot 0]}$$

It is evident that $[A_1A_1\cdot 0] = (A_1, A_1) > 0$, since the matrix (6.1) is properly positive. By Lemma 7.3) the value of $D(\zeta)$ and the determinant of (7.5) are equal, and furthermore the minor of the element $[A_1A_1\cdot 0]$ is a symmetric matrix of properly positive type. Thus $[A_2A_2\cdot 1] > 0$, and $[A_tA_s\cdot 1] = [A_sA_t\cdot 1]$.

The minor of $[A_1A_1\cdot 0]$ surely satisfies all the conditions in Lemma 7.3). We may, therefore, apply a transformation of the form (7.2) to the minor of $[A_1A_1\cdot 0]$, and secure another matrix of the same character as (7.5). In other words, we may multiply on the left of the matrix (7.5) by

$$\text{(7.6)} \qquad \tau_2 = \delta + F_2D_2$$

where D_2 is the second row of the r row identity matrix δ, and

$$F_2(n) = 0 \quad (n \leqq 2); \qquad F_2(n) = \frac{[A_2A_n\cdot 1]}{[A_2A_2\cdot 1]} \qquad (n > 2)\,.$$

In general, let

$$\text{(7.7)} \qquad \tau_i = \delta + F_iD_i \qquad (i = 1, \cdot\cdot, r-1)\,,$$

where D_i is the i^{th} row of the r row identity matrix δ, and

$$\text{(7.8)} \qquad F_i(n) = 0 \quad (n \leqq i)\,; \qquad F_i(n) = -\frac{[A_iA_n\cdot i-1]}{[A_iA_i\cdot i-1]} \qquad (n > i)\,.$$

Continuous application of this type of transformation ultimately reduces the matrix (6.1) to the form

$$(7.9)\qquad \eta = \begin{matrix} [A_1A_1\cdot 0] & [A_1A_2\cdot 0] & [A_1A_3\cdot 0] & \cdots & [A_1A_r\cdot 0] \\ 0 & [A_2A_2\cdot 1] & [A_2A_3\cdot 1] & \cdots & [A_2A_r\cdot 1] \\ 0 & 0 & [A_3A_3\cdot 2] & \cdots & [A_3A_r\cdot 2] \\ \\ 0 & 0 & 0 & & [A_rA_r\cdot r-1] \end{matrix}$$

where

$$(7.9_1)\qquad [A_tA_s\cdot h] = \frac{[A_hA_h\cdot h-1][A_tA_s\cdot h-1] - [A_hA_t\cdot h-1][A_hA_s\cdot h\ 1]}{[A_hA_h\cdot h-1]}$$

$$(t, s = 1, \cdots, r; \qquad 0 \leqq h \leqq sm(t, s)).^{14}$$

In the matrix (7.9), we see by virtue of Lemma 7.3) that $[A_iA_i\cdot i-1] > 0$ for every i, and $[A_tA_s\cdot h] = [A_sA_t\cdot h]$ for every s, t and $0 \leqq h \leqq sm(t, s)$. If $h = sm(t, s)$, then $[A_tA_s\cdot h] = 0$.

Let $\tau = \tau_{r-1}\cdot\tau_{r-2}\cdots\tau_1$. Then by the associative law of multiplication of matrices, we see that

$$(7.10)\qquad \eta = (\tau_{r-1}\cdots\tau_1)\zeta = \tau\zeta.$$

Thus we prove

THEOREM 15. *If the set of vectors $A_1, \cdots, A_r$ is linearly independent, then there exists a square matrix τ of order such that $\tau\zeta$ is of the form* (7.9) *where all elements below the principal diagonal are* 0; *every element in the principal diagonal* $[A_iA_i\cdot i-1]$ $(i = 1, \cdots, r)$, *is greater than zero; and* $[A_tA_s\cdot h] = [A_sA_t\cdot h]$ *for* $s, t = 1, \cdots, r$, *and* $h < sm(t, s)$. *Furthermore the determinants of the matrices* (6.1) *and* (7.9) *are equal.*

We now prove the following lemma which will be useful in the later section.

LEMMA 7.4). *If* $[A_iA_i\cdot i-1]$ *is different from zero for every* $i \geqq 0$, *then for every pair of integers* (s, t), *where* $s, t = 1, \cdots, r$, *and* $n \leqq sm(t, s)$, *we have*

$$\text{a)}\qquad [A_tA_s\cdot n] = (A_t, A_s) - \sum_{i=1}^{n-1}\frac{[A_iA_t\cdot i-1]}{[A_iA_i\cdot i-1]}[A_iA_s\cdot i-1].$$

$$\text{b)}\qquad [A_t(A_s + A_u)\cdot n] = [A_tA_s\cdot n] + [A_tA_u\cdot n], \quad (u = 1, \cdots, r).$$

$$\text{c)}\qquad [(cA_t)A_s\cdot n] = c[A_tA_s\cdot n], \qquad (c = a \text{ constant}).$$

To prove a), take every pair (s, t). We find the lemma is true for $n = 0$. Assuming it is true for every (s, t) and for $n = h < sm(s, t)$, we find that $h + 1 \leqq sm(s, t)$, and

$$(A_t, A_s) - \sum_{i=1}^{h+1}\frac{[A_iA_t\cdot i-1]}{[A_iA_i\cdot i-1]}[A_iA_s\cdot i-1]$$

[14] sm (s, t) read "the smaller one of (t, s)."

$$= (A_t, A_s) - \sum_{i=1}^{h} \frac{[A_iA_t \cdot i-1]}{[A_iA_i \cdot i-1]} [A_iA_s \cdot i-1] - \frac{[A_{h+1}A_t \cdot h]}{[A_hA_h \cdot h]} [A_{h+1}A_s \cdot h]$$

$$= [A_tA_s \cdot h] - \frac{[A_{h+1}A_t \cdot h]}{[A_hA_h \cdot h]} [A_{h+1}A_s \cdot h] = [A_tA_s \cdot h+1],$$

for every s, t.

Parts b) and c) are true for $n = 0$. Now make use of the equality in a) and prove by induction.

VIII. Gauss's Method of Substitution and its Relation to Gramian Schmidt's Orthogonalization Process

Let us write the set of observations in the form:

$$\begin{pmatrix} a_{11} & a_{12} & & a_{1n} \\ \\ a_{r1} & a_{r2} & \cdots & a_{rn} \end{pmatrix}$$

Let the orthogonalized set also be written in the form

$$\begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \\ b_{r1} & & b_{rn} \end{pmatrix}$$

From Theorems 5 and 6, we find that there exists a transformation κ given by an r-row square matrix such that $\beta = \kappa\alpha$. Thus by the associative law of multiplication of matrices, we have

$$\beta\alpha' = (\kappa\alpha)\alpha' = \kappa(\alpha\alpha').$$

Now the matrix $\alpha\alpha'$ is the Gramian matrix (6.1). Thus

$$\beta\alpha' = \kappa\zeta. \tag{8.1}$$

The composite matrix $\beta\alpha'$ is of the form

$$\begin{bmatrix} (B_1, A_1)(B_1, A_2) \cdots (B_1, A_r) \\ (B_2, A_1)(B_2, A_2) \cdots (B_2, A_r) \\ \\ (B_r, A_1)(B_r, A_2) \cdots (B_r, A_r) \end{bmatrix} \tag{8.2}$$

By Theorems 5 and 6, we note that $(B_s, A_t) = 0$ for $s > t$, and $(B_s, A_s) = (B_s, B_s)$ for every s. Thus the preceding matrix can be written in the form

$$\begin{bmatrix} (B_1, B_1)(B_1, A_2)(B_1, A_3) \cdots (B_1, A_r) \\ 0 \quad (B_2, B_2)(B_2, A_3) \cdots (B_2, A_r) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ 0 \quad 0 \quad 0 \quad \cdots (B_r, B_r) \end{bmatrix} \tag{8.3}$$

We have proved the following theorem:

THEOREM 16. *Let $A_1, \cdots, A_r$ be a set of vectors, and $B_1, \cdots, B_r$ be the orthogonalized set; and let $\alpha = (a_{ij})$, $\beta = (b_{ij})$. Then there exists a square r-row matrix κ such that $\beta = \kappa\alpha$, and $\kappa\alpha\alpha'$ is a matrix of the form* (8.3) *where all the elements below the principal diagonal are zeros and every element in the principal diagonal is positive. If the set $A_1, \cdots, A_r$ is linearly independent, then every element in the principal diagonal is greater than zero.*

THEOREM 17. *Let $A_1, \cdots, A_r$ be a set of vectors and $B_1, \cdots, B_r$ be the orthogonalized set; and let $\alpha = (a_{ij})$, $\beta = (b_{ij})$. Then $D(\beta\alpha') = D(\alpha\alpha')$.*

For by equations (2.1), we note that $D(\kappa) = 1$. Thus

$$D(\beta\alpha') = D(\kappa\alpha\alpha') = D(\kappa)D(\alpha\alpha') = D(\alpha\alpha') .$$

THEOREM 18. *If the set of vectors, $A_1, \cdots, A_r$ is linearly independent, the matrix κ arising from Gram-Schmidt's orthogonalization process is identical with the matrix τ defined by* (7.10).

To prove this theorem, we first establish the following

LEMMA 8.5): *If the set $A_1, \cdots, A_r$ be linearly independent, and $B_1, \cdots, B_r$ be the orthogonalized set, then for every t, h, we have*

$$(B_h, A_t) = [A_h A_t \cdot h - 1] .$$

By Theorem 10, the set B_i is linearly independent, and hence $n(B_i) > 0$ for every i. The lemma is evidently true for every t and $h = 1$. Assuming it is true for every t and $h = s$, we shall prove it is also true for every t and $h = s - 1$. Now

$$B_{s+1} = A_{s+1} - \sum_{i=1}^{s} \frac{(A_s, B_i)}{(B_i, B_i)} B_i = A_{s+1} - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} B_i .$$

Thus by the linear property of (,) we secure, for every t

$$\begin{aligned}(B_{s+1}, A_t) &= \left(A_{s+1} - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} B_i, A_t\right) \\ &= (A_{s+1}, A_t) - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} (B_i, A_t) \\ &= (A_{s+1}, A_t) - \sum_{i=1}^{s} \frac{[A_i A_s \cdot i - 1]}{[A_i A_i \cdot i - 1]} [A_i A_t \cdot i - 1] \\ &= [A_{s+1} A_t \cdot s]\end{aligned}$$

by virtue of lemma 4.4).

From this lemma, we conclude at once that the matrices (7.9) and (8.3) are equal. Thus by (8.1), we have

$$\kappa\zeta = \beta\alpha' = \tau\zeta, \quad \text{or} \quad (\kappa - \tau)\zeta = \omega .$$

Since ζ is non-singular (by Theorem 12), we have

$$\omega = (\kappa - \tau)\zeta\zeta^{-1} = (\kappa - \tau)\delta = \kappa - \tau,$$

which proves the theorem.

From Lemma 8.5), we have

Lemma 8.6). Let $L = (l_1, \cdots, l_n)$. Suppose the set $A_1, \cdots, A_r$ to be linearly independent, and $B_1, \cdots, B_r$ to be the orthogonalized set. Then for every h,

$$[A_h L \cdot h - 1] = (B_h, L) .$$

Theorems 16, 17, and 18 furnish us a new method for finding the most probable values of the unknowns in the theory of least squares. The formulation of the system of normal equations may be omitted in this new procedure, which may be described briefly as follows: After we obtain a set of observations $A_1, \cdots, A_r$, we orthogonalize this set by means of Gram-Schmidt's process. Let L be a non-zero vector. The product

$$\begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \\ \\ b_{r1} & \cdots & b_{rn} \end{pmatrix} \quad \begin{pmatrix} a_{11} & \cdots & a_{r1}, & -l_1 \\ \\ \\ a_{1n} & \cdots & a_{rn}, & -l_n \end{pmatrix}$$

will give us the result as desired by Gauss's method of substitution.

Academia Sinica,
Peiping, China.

A NOTE ON THE ANALYSIS OF VARIANCE[1]

By Solomon Kullback

By considering a set of independent items classified in some relevant manner into N sets of s items each, and by the use of a dispersion theorem of Prof. J. L. Coolidge,[2] Prof. H. L. Rietz[3] arrives at estimates of variance, used by Dr. R. A. Fisher, without making use of arguments involving the number of degrees of freedom of the items concerned.

By proceeding along the lines followed by Coolidge and Rietz but considering a set of independent items classified into N sets of $s_i (i = 1, 2, \cdots, N)$ items each, we shall arrive at certain other important results of R. A. Fisher[4] in his analysis of variance.

The theorem referred to above is as follows: If n independent quantities $y_1, y_2, \cdots, y_n$ be given, their expected values being $a_1, a_2, \cdots, a_n$, while the expected values of their squares are $A_1, A_2, \cdots, A_n$, respectively, and if we agree to set $y = (1/n) \sum_{i=1}^{n} y_i$, $a = (1/n) \sum_{i=1}^{n} a_i$, then the expected value of the variance, $(1/n) \sum_{i=1}^{n} (y_i - y)^2$ is

$$\frac{1}{n}\left[\frac{n-1}{n} \sum_{i=1}^{n} (A_i - a_i^2) + \sum_{i=1}^{n} (a_i - a)^2\right]. \tag{1}$$

Suppose a set of independent items has been classified in some relevant manner into N sets of s_i $(i = 1, 2, \cdots, N)$ items each as follows:

$$\begin{array}{cccccc}
x_{11}, & x_{12}, & \cdots, & x_{1s_1}, & \bar{x}_1 \\
x_{21}, & x_{22}, & \cdots, & x_{2s_2}, & \bar{x}_2 \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
x_{N1}, & x_{N2}, & \cdots, & x_{Ns_N}, & \bar{x}_N \\
 & & & & \bar{x}
\end{array} \tag{2}$$

where $\bar{x}_i (i = 1, 2, \cdots, N)$ is the arithmetic mean of the i^{th} set and $\bar{x}$ the mean of the pooled sample of $s = s_1 + s_2 + \cdots + s_N$ items.

We shall assume that the set (2) is statistically homogeneous in the sense that,

[1] Presented to the American Mathematical Society, February 23, 1935.

[2] Bulletin Am. Math. Soc., Vol. 27 (1921) p. 439.

[3] Bulletin Am. Math. Soc., Vol. 38 (1932) pp. 731–735.

[4] Proceedings of the International Math. Congress, Toronto, 1924, Vol. 2, p. 802 ff.

using $E(\quad)$ for the expected value of the expression in the parenthesis, we may let $E(x_{ij}) = a$, $E(x_{ij}^2) = A$, $(i = 1, 2, \cdots, N, j = 1, 2, \cdots, s_i)$. Then, using (1)

$$E\left(\sum_{j=1}^{s_i} (x_{ij} - \bar{x}_i)^2\right) = (s_i - 1)(A - a^2). \tag{3}$$

Summing (3) from $i = 1$ to N, we have

$$E\left(\sum_{i=1, j=1}^{N, s_i} (x_{ij} - \bar{x}_i)^2\right) = (A - a^2)\sum_{i=1}^{N}(s_i - 1) = (s - N)(A - a^2). \tag{4}$$

Similarly, by using (1)

$$E\left(\sum_{i=1}^{N} s_i(\bar{x}_i - \bar{x})^2\right) = \frac{N-1}{N}\sum_{i=1}^{N} s_i\,[E(\bar{x}_i^2) - a^2]. \tag{5}$$

But[5]

$$E(\bar{x}_i^2) - a^2 = E(\bar{x}_i - a)^2, \qquad \text{and} \tag{6}$$

$$E(\bar{x}_i - a)^2 = (A - a^2)/s_i, \qquad \text{therefore} \tag{7}$$

$$E\left(\sum_{i=1}^{N} s_i(\bar{x}_i - \bar{x})^2\right) = (N - 1)(A - a^2). \tag{8}$$

Similarly by using (1)

$$E\left(\sum_{i=1, j=1}^{N, s_i} (x_{ij} - \bar{x})^2\right) = (s - 1)(A - a^2). \tag{9}$$

Thus, in a statistically homogeneous set of items, classified as in (2), the following estimates of Variance have the same expected value:

$$\begin{aligned}
V &= \frac{S}{s-1}, &\quad \text{where} \quad S &= \sum_{i=1, j=1}^{N, s_i} (x_{ij} - \bar{x})^2 \\
V_i &= \frac{S_i}{s-N}, &\quad \text{where} \quad S_i &= \sum_{i=1, j=1}^{N, s_i} (x_{ij} - \bar{x}_i)^2 \\
V_{\bar{x}} &= \frac{S_{\bar{x}}}{N-1}, &\quad \text{where} \quad S_{\bar{x}} &= \sum_{i=1}^{N} s_i(\bar{x}_i - \bar{x})^2.
\end{aligned} \tag{10}$$

These estimates are used in applying the analysis of variance to the study of the correlation ratio, η, for uncorrelated material, where $\eta^2 = S_{\bar{x}}/S$.

OFFICE OF THE CHIEF SIGNAL OFFICER,
WASHINGTON, D. C.

[5] Rietz, H. L., loc. cit. p. 733.

A PROBLEM INVOLVING THE LEXIS THEORY OF DISPERSION

By Walter A. Hendricks

The attention of the author was recently directed to a study of the hatchability of chicken eggs at the U. S. Animal Husbandry Experiment Station, Beltsville, Maryland. It was necessary to find the average hatchability of the fertile eggs incubated for each of a number of lots of birds and the corresponding standard errors of those averages.

It was very apparent that some methods for computing such values, in common use at the present time, do not give satisfactory results. This is due to the fact that the fertile eggs produced by different birds vary considerably with respect to hatchability as well as with respect to number of eggs available for incubation. It seems reasonable to suppose that the variability in hatchability of a number of fertile eggs, produced by a given number of birds, should obey the Lexis law of dispersion. This supposition is based on two hypotheses:

(a) The probability that a fertile egg will hatch is constant for all fertile eggs produced by the same bird during the time interval under consideration.

(b) The probability that a fertile egg will hatch varies from bird to bird.

The reader familiar with the principles of genetics may question the validity of the first of these hypotheses. The probability that a fertile egg will hatch is largely governed by the genes carried by the chromosomes of the ovum of the hen and the sperm of the male bird which fertilized that ovum. The kinds of genes carried by various ova and spermatozoa are not necessarily the same, even when those ova and spermatozoa are produced by the same female and male birds, respectively. However, if we have a sample of a number of fertile eggs produced by the same hen, we are justified in assuming that the proportion of those eggs which will hatch is constant, except for sampling fluctuations, when successive samples of fertile eggs produced by the given hen are incubated, provided, of course, that the eggs in the successive samples were fertilized by the same male bird or birds. The limit approached by the proportion of fertile eggs which hatch as the number of fertile eggs produced by the given hen becomes infinitely large may be defined as the probability that a fertile egg produced by that hen will hatch. It will be recognized that this definition is based on purely academic considerations, since there are physical limitations to the number of fertile eggs which a hen can produce in a given period of time. Hypotheses (a) and (b) are to be interpreted in the light of this definition of the probability that a fertile egg produced by a given bird will hatch.

Let $s_1, s_2, \cdots s_n$ represent the numbers of fertile eggs produced by n birds during a period of time and let $f_1, f_2, \cdots f_n$, respectively, represent the numbers

of chicks obtained from those eggs when the eggs are incubated. Let $p_k = \frac{f_k}{s_k}$ represent the hatchability of the fertile eggs produced by the k^{th} bird.

The squared standard error of p_k is given by the Lexis formula:[1]

$$\sigma^2_{p_k} = \frac{PQ}{s_k} + \frac{s_k - 1}{n s_k} \sum_{t=1}^{n} (P_t - P)^2 \tag{1}$$

in which the P_t represent the respective probabilities that the fertile eggs produced by the n birds will hatch, P is the arithmetic mean of the P_t, and Q is equal to $1 - P$.

The values of the probabilities, P_t, are not known. However, as a first approximation to equation (1) we may write:

$$\sigma^2_{p_k} = \frac{pq}{s_k} + \frac{s_k - 1}{n s_k} \sum_{t=1}^{n} (p_t - p)^2 \tag{2}$$

in which p is the arithmetic mean of the p_t and q is equal to $1 - p$.

The product, pq, can be accepted as a reasonably close approximation to the product, PQ, but the expression, $\sum_{t=1}^{n} (p_t - p)^2$, will, in general, be greater than the expression, $\sum_{t=1}^{n} (P_t - P)^2$. The reason for this is apparent when we consider that if each of these two expressions is divided by n, the former yields an estimate of the squared standard deviation of the p_t while the latter yields an estimate of the squared standard deviation of the P_t. The standard deviation of the p_t will, in general, be greater than that of the P_t because the p_t are more or less imperfect estimates of the P_t and are, therefore, subject to sampling errors from which the P_t are free.

We may write:

$$\frac{1}{n} \sum_{t=1}^{n} (P_t - P)^2 = \frac{1}{n} \sum_{t=1}^{n} (p_t - p)^2 - \sigma^2_c \tag{3}$$

in which σ^2_c is an appropriate correction as yet undefined.

Since the p_t would approach the P_t as statistical limits if each of the s_t were made extremely large, it follows that σ^2_c must approach zero as each of the s_t approaches infinity. Furthermore, if $P_1 = P_2 = \cdots P_n = P$, we must have:

$$\frac{1}{n} \sum^{n} (p_t - p)^2 - \sigma^2_c = 0 \qquad \text{or}$$

$$\sigma^2_c = \frac{1}{n} \sum_{t=1}^{n} (p_t - p)^2 . \tag{4}$$

[1] The formula as given in this paper is a modification of that given by Rietz, H. L. (1927) in his book, Mathematical Statistics, Open Court Publishing Co., Chicago, which was necessary in order to make it applicable to relative frequencies.

These conditions suggest that σ_c^2 be defined by the equation:

$$\sigma_c^2 = \frac{pq}{n} \sum_{t=1}^{n} \frac{1}{s_t}. \tag{5}$$

If σ_c^2 is so defined, it will obviously approach zero as each of the s_t approaches infinity. Furthermore, it has been shown by Yule[2] that if we have a series of n relative frequencies, such as the p_t under discussion, based on n samples of unequal size, and the probabilities of the occurrence and non-occurrence, respectively, of the particular event under consideration are constant from sample to sample, the squared standard deviation of those relative frequencies is given by a relation such as that used to define σ_c^2 in equation (5). Therefore, the second condition is also satisfied. σ_c^2 may be interpreted as representing that part of the squared standard deviation of the p_t which is due to the unreliability of the p_t as estimates of the P_t.

Therefore, it seems reasonable to write:

$$\frac{1}{n} \sum_{t=1}^{n} (P_t - P)^2 = \frac{1}{n} \sum_{t=1}^{n} (p_t - p)^2 - \frac{pq}{n} \sum_{t=1}^{n} \frac{1}{s_t}. \tag{6}$$

Combining equations (1) and (6), we obtain the following formula for calculating the squared standard error of p_k:

$$\sigma_{p\,k}^2 = \frac{pq}{s_k} + \frac{s_k - 1}{ns_k} \left[\sum_{t=1}^{n} (p_t - p)^2 - pq \sum_{t=1}^{n} \frac{1}{s_t} \right].$$

Since the weight of a measurement is inversely proportional to the square of the standard error of the measurement, we are now in a position to calculate a weighted mean, $\bar{p}$, of the p_t.

$$\bar{p} = \frac{\sum_{t=1}^{n} w_t p_t}{\sum_{t=1}^{n} w_t} \tag{8}$$

in which:

$$w_t = \frac{1}{\sigma_{p_t}^2}. \tag{9}$$

The squared standard error of $\bar{p}$ is given by the familiar formula:

$$\sigma_{\bar{p}}^2 = \frac{\sum_{t=1}^{n} w_t (p_t - \bar{p})^2}{(n-1) \sum_{t=1}^{n} w_t}. \tag{10}$$

[2] Yule, G. Udny, 1927. Introduction to the Theory of Statistics, Charles Griffin and Co., London.

It would seem that $\bar{p}$ may be accepted as a good estimate of the average hatchability of the fertile eggs produced by the given lot of birds, and that equation (10) may be used to obtain a valid estimate of the reliability of $\bar{p}$.

However, the problem is not quite so simple. In the first place, there is usually a small amount of positive correlation between the number of fertile eggs produced by a bird and the hatchability of those eggs. Secondly, as pointed out earlier in this paper, the hatchability of fertile eggs is influenced to some extent by the male birds used to fertilize the eggs. The error involved in neglecting the correlation between hatchability and number of fertile eggs incubated does not seem to be of much importance in those practical problems which have come to the author's attention. The effects of differences among the male birds may be largely obviated in experimental work by frequently transferring male birds from lot to lot during the experimental period.

The best test of the suitability of a particular formula for calculating the standard error of an average is to compare the value of the standard error calculated by means of the formula with the corresponding value obtained by direct calculation from the distribution of a number of such averages obtained under essentially the same conditions. The accompanying table gives the standard error of the weighted average hatchability of fertile eggs calculated for each of four lots of birds by means of equation (10), together with the corresponding values obtained from the distribution of averages. The former are designated as the "predicted" values and the latter are designated as the "observed" values. In the calculation of the "observed" values, the various averages were assigned the same weights which were used in the calculation of the "predicted" values.

Comparison of "predicted" and "observed" standard errors of the weighted average hatchability of fertile eggs, calculated for each of four lots of birds

Lot	$\bar{p}$	Standard error of $\bar{p}$	
		"Predicted"	"Observed"
1	0.7684	0.0287	0.0327
2	0.7115	0.0533	0.0561
3	0.6834	0.0355	0.0379
4	0.7260	0.0615	0.0674

The data used in these calculations involved a total of 74 birds, approximately equally divided among the four lots, and a total of 2,901 fertile eggs which were produced and incubated during an experimental period of 48 weeks. The agreement between the "predicted" and "observed" standard errors of the weighted average hatchability for each lot of birds is excellent. However, the author's experience with biological data tends to make him doubt that such

close agreement will always be found when such data are subjected to the above treatment. The agreement in the present illustration could be less close without indicating that the method of calculating the "predicted" standard errors is unsound.

BUREAU OF ANIMAL INDUSTRY,
U. S. DEPARTMENT OF AGRICULTURE,
WASHINGTON, D. C.

A METHOD FOR DETERMINING THE COEFFICIENTS OF A CHARACTERISTIC EQUATION

PAUL HORST

For the characteristic equation

$$\begin{vmatrix} a_{11} - x & \cdots & a_{1n} \\ & & \\ a_{n1} & & \end{vmatrix} \equiv (-1)^n(x^n - c_1x^{n-1} + c_2x^{n-2} - \cdots + c_n) \equiv (x - \alpha_1)(x - \alpha_2) \cdots (x - \alpha_n) \tag{1}$$

it is well known that

$$c_i = A_i$$

where A_i is the sum of all i^{th} order co-axial minors of the determinant

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ & & \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} \tag{2}$$

If n exceeds 3 or 4, the process of calculating all possible principal minors is very cumbersome.

But another more systematic method of calculating the c's may be adopted. Suppose we define

$$a_{n1}^{(p)} \cdots a_{nn}^{(p)} \tag{3}$$

and

$$\sum^{n} \alpha_i^{(p)} = S_p\,. \tag{4}$$

It may be proved[1] that

$$S_p = \sum^{n} a_{ii}^{(p)}\,. \tag{5}$$

But from Newton's identities[2] we have

$$S_p + c_1S_{p-1} + c_2S_{p-2} + \cdots + c_{p-1}S_1 + pc_p = 0\,. \tag{6}$$

[1] Muir, L. & Metzler, W. H., "A Treatise on the Theory of Determinants," p. 606, ¶ 650 and 651.

[2] Dickson, L. E., "First Course in the Theory of Equations," p. 134, ¶ 106.

Newton's identities are ordinarily employed for calculating the sums of the powers of the roots of a polynomial when the coefficients are known. They may be employed equally well, however, for calculating the coefficients when the sums of the powers are given. Thus by means of equations (5) and (6) the coefficients of (1) may be readily calculated.

If in (2) $a_{ij} = a_{ji}$, the calculation of the successive A^p values is straightforward. The determinant A is used as a constant multiplier so that

$$A \cdot A = A^2, \qquad A \cdot A^2 = A^3, \cdots A \cdot A^{n-1} = A^n$$

and the multiplication is column by column. That is,

$$a_{ij}^{(1+p)} = \sum_{k=1}^{n} a_{ki} a_{kj}^{(p)} .$$

THE GENERALIZED PROBLEM OF CORRECT MATCHINGS

By Dwight W. Chapman

A method common to many experimental and testing procedures in psychology and education is to require an individual to match, as best he can, members of one series of items with members of a second series of quite different items certain of which are in some sense true apposites of items in the first series. Thus the experimental psychology of personality has often investigated the ability of graphologists or laymen to pair samples of handwriting produced by a group of persons with, say, character-sketches of these same persons; and the excess of correct matchings thus produced over the number to be expected by chance has been used as evidence that the expressive movement of handwriting affords characteristics diagnostic of personal traits. Fortunately, the excesses experimentally obtained have often been so large as obviously to exclude the operation of chance alone. But many empirical results show small excesses only; and the interpretation of such findings has not hitherto been subjected to rigid statistical analysis.

The particular statistical problem resident in this experimental procedure is twofold, involving the estimation of the significance of (a) a given number of correct matchings produced by one individual, and (b) a given mean number of correct matchings produced by a group of individuals working with the same material independently.

Furthermore, two cases arise in practice: (1) the two series of items are of equal length, and each item in either series has a true apposite in the other series; or (2) the two series may be of unequal length, in which case the longer series contains not only a true apposite for each item of the shorter series, but, in addition, a certain number of extra, irrelevant items which cannot be correctly matched with any items in the shorter series. I have already given the solution to problems (a) and (b) for case (1).[1] But case (1) forms only a corollary of the more general case (2), to the solution of which this present paper is devoted.

(a) The Significance of a Given Number of Correct Matchings Resulting from a Single Trial

Let there be given a series of u x-items,

$$x_1, x_2, \cdots x_t, \cdots x_u$$

and a series of t y-items,

$$y_1, y_2, \cdots y_t .$$

[1] The Statistics of the Method of Correct Matchings, *Amer. Jour. Psychol.*, 46, 1934, 287–298.

Let $t \leqq u$, and let the first t x-items be in some sense true apposites of the correspondingly numbered y-items, so that if y_j be paired with x_j $(j = 1, 2, \cdots t)$, this pairing will constitute a correct matching.

The first problem which arises is that of determining the probability that a single random arrangement of the t y-items against t of the x-items will result in exactly s $(= 0, 1, 2, \cdots t)$ correct matchings.

We begin by putting the first s y-items in correspondence with their apposite x-items. Then the number of arrangements of the t y-items in which only these s are correctly matched is the number of arrangements of the remaining $t - s$ y-items against the remaining $u - s$ x-items such that no correct matchings occur. With respect to these items, let

$n =$ the number of all possible arrangements,

$n(Y_j) =$ the number of arrangements such that at least the j^{th} item is correctly matched with its apposite,

$n(Y_jY_k) =$ the number of arrangements such that at least both the j^{th} and k^{th} items are matched with their apposites, etc.;

and let

$n(\bar{Y}_j) =$ the number of arrangements such that at least the j^{th} item is not matched with its apposite,

$n(\bar{Y}_j\bar{Y}_k) =$ the number of arrangements such that at least the j^{th} and k^{th} items are not matched with their apposites, etc.

We have then to evaluate the expression $n(\bar{Y}_{s+1}\bar{Y}_{s+2} \cdots \bar{Y}_t)$, the number of arrangements of the items remaining, after setting s of them correctly matched, such that no further correct matchings occur.

Now it can be shown that[2]

$$
\begin{aligned}
n(\bar{Y}_{s+1}\bar{Y}_{s+2} \cdots \bar{Y}_t) = n & \\
& - [n(Y_{s+1}) + n(Y_{s+2}) + \cdots + n(Y_t)] \\
& + [n(Y_{s+1}Y_{s+2}) + n(Y_{s+1}Y_{s+3}) + \cdots + n(Y_{t-1}Y_t)] \\
& - [n(Y_{s+1}Y_{s+2}Y_{s+3}) + \cdots + n(Y_{t-2}Y_{t-1}Y_t)] \\
& + \cdots \\
& +(-1)^t\, n(Y_{s+1}Y_{s+2} \cdots Y_t)\,.
\end{aligned}
$$

The value of the expressions on the right side of this equation can be determined as follows:

[2] H. Whitney, A Logical Expansion in Mathematics, *Bull. Amer. Math. Soc.*, 1932, 572–579.

The value of n is the number of ways in which $t - s$ items can be arranged against

$u - s$ items, which is $\dfrac{(u-s)!}{[(u-s)-(t-s)]!} = \dfrac{(u-s)!}{(u-t)!}$.

The value of the first bracket—the number of arrangements of these items such that some one of them is correctly matched—is derived by holding one of the items matched, which can be chosen in $t - s$ ways. This leaves $t - s - 1$ y-items, which can be arranged against the remaining $u - s - 1$ x-items in $(u - s - 1)!/(u - t)!$ ways. The product of these two expressions gives us for the value of the first bracket

$$[n(Y_{s+1}) + \cdots + n(Y_t)] = \frac{(t-s)!(u-s-1)!}{(u-t)!}.$$

To evaluate the second bracket, we hold two of the $t - s$ items matched, which can be chosen in $(t - s)!/[2!(t - s - 2)!]$ ways. There remains $t - s - 2$ y-items which can be arranged against the remaining $u - s - 2$ x-items in $(u-s-2)!/(u-t)!$ ways. The product of these two expressions gives us

$$[n(Y_{s+1}Y_{s+2}) + \cdots + n(Y_{t-1}Y_t)] = \frac{(t-s)!(u-s-2)!}{2!(t-s-2)!(u-t)!}.$$

Continuing thus, we develop the following series for the number of arrangements of t items against u items such that the first s are correctly matched:

$$n(\bar{Y}_{s+1}\bar{Y}_{s+2}\cdots\bar{Y}_t) = \frac{(u-s)!}{(u-t)!} - \frac{(t-s)!(u-s-1)!}{(u-t)!} + \frac{(t-s)!(u-s-2)!}{2!(t-s-2)!(u-t)!} - \cdots + (-1)^{t-s}\frac{(t-s)!(u-t)!}{(t-s)!(u-t)!}.$$

In order to express the number of arrangements, $N_{(s)}$, such that *any* s correct matchings occur, we must multiply the above series by $t!/[s!(t - s)!]$, which is the number of ways in which s items can be chosen from t items:

$$N_{(s)} = \frac{t!}{s!(t-s)!}\left[\frac{(u-s)!}{(u-t)!} - \frac{(t-s)!(u-s-1)!}{(u-t)!} + \cdots + (-1)^{t-s}\frac{(t-s)!(u-t)!}{(t-s)!(u-t)!}\right].$$

And in order to obtain the probability that a single random arrangement will result in exactly s correct matchings, we must further divide by $u!/(u - t!)$, which is the total number of ways in which t items can be arranged against u items. Calling this probability $P_{(s)}$, we have then

$$P_{(s)} = \frac{t!\,(u-t)!}{u!s!\,(t-s)!}\left[\frac{(u-s)!}{(u-t)!} - \frac{(t-s)!\,(u-s-1)!}{(u-t)!} + \cdots + (-1)^{t-s}\frac{(t-s)!\,(u-t)!}{(t-s)!\,(u-t)!}\right].$$

Finally, factoring $(t-s)!/(u-t)!$ out of all terms in the bracket, the series simplifies to[3]

$$P_{(s)} = \frac{t!}{s!u!}\left[\frac{(u-s)!}{0!(t-s)!} - \frac{(u-s-1)!}{1!(t-s-1)!} + \frac{(u-s-2)!}{2!(t-s-2)!} - \cdots + (-1)^{t-s}\frac{(u-t)!}{(t-s)!\,0!}\right]. \qquad (1)$$

In any practical situation, the significant question is not the probability that exactly s correct matchings shall occur, but the probability of s *or more* correct matchings. Obviously

$$P_{(s\text{ or more})} = P_{(s)} + P_{(s+1)} + \cdots + P_{(t)}\,.$$

whence, by equation (1),

$$\begin{aligned} P_{(s\text{ or more})} = {} & \frac{t!}{s!u!}\left[\frac{(u-s)!}{0!(t-s)!} - \frac{(u-s-1)!}{1!(t-s-1)!} + \frac{(u-s-2)!}{2!(t-s-2)!} - \cdots + (-1)^{t-s}\frac{(u-t)!}{(t-s)!0!}\right] \\ & + \frac{t!}{(s+1)!u!}\left[\frac{(u-s-1)!}{0!(t-s-1)!} - \frac{(u-s-2)!}{1!(t-s-2)!} + \cdots + (-1)^{t-s-1}\frac{(u-t)!}{(t-s-1)!0!}\right] \\ & + \frac{t!}{(s+2)!u!}\left[\frac{(u-s-2)!}{0!(t-s-2)!} - \cdots + (-1)^{t-s-2}\frac{(u-t)!}{(t-s-2)!0!}\right] \\ & + \cdots \\ & + \frac{t!}{t!u!}\left[\frac{(u-t)!}{0!0!}\right]. \end{aligned} \qquad (2)$$

Or, collecting terms in a form better suited to practical computation from tables of factorials and reciprocals,

$$\begin{aligned} P_{(s\text{ or more})} = \frac{t!}{u!}\Bigg\{ & \frac{(u-s)!}{(t-s)!}\left[\frac{1}{0!s!}\right] \\ & + \frac{(u-s-1)!}{(t-s-1)!}\left[\frac{1}{0!(s+1)!} - \frac{1}{1!s!}\right] \\ & + \frac{(u-s-2)!}{(t-s-2)!}\left[\frac{1}{0!(s+2)!} - \frac{1}{1!(s+1)!} + \frac{1}{2!s!}\right] \end{aligned}$$

[3] In the special case in which the series of x-items and the series of y-items are of the same length, whence $t = u$, equation (1) reduces to

$$P_{(s)} = \frac{1}{s!}\left[\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \cdots + (-1)^{t-s}\frac{1}{(t-s)!}\right].$$

$$+ \cdots + \frac{(u-t)!}{0!}\left[\frac{1}{0!t!} - \frac{1}{1!(t-1)!} + \frac{1}{2!(t-2)!} \cdots + (-1)^{t-s}\frac{1}{(t-s)!s!}\right]\Bigg\}.$$

(b) The Significance of a Given Mean Number of Correct Matchings Resulting from n Independent Trials

A frequent practical situation is that in which interest centers on the significance of the mean number of correct matchings achieved by a group of n individuals working independently with the same two series.

In order to determine the probability that the mean number of correct matchings, $\bar{s}$, resulting from n independent trials shall equal or exceed a given value, we are required to describe the distribution of the means of samples of size n drawn at random from a parent population in which the variable is $s(= 0, 1, 2, \cdots t)$ with relative frequencies $P_{(0)}, P_{(1)}, P_{(2)}, \cdots P_{(t)}$, given by equation (1). The tabulation of this parent distribution follows:

Table I: Distribution of s

s	*Relative frequency* $(= P_{(s)})$
0	$\frac{t!}{0!u!}\left[\frac{u!}{0!t!} - \frac{(u-1)!}{1!(t-1)!} + \frac{(u-2)!}{2!(t-2)!} - \frac{(u-3)!}{3!(t-3)!} + \cdots + (-1)^t\frac{(u-t)!}{t!0!}\right]$
1	$\frac{t!}{1!u!}\left[\frac{(u-1)!}{0!(t-1)!} - \frac{(u-2)!}{1!(t-2)!} + \frac{(u-3)!}{2!(t-3)!} - \cdots + (-1)^{t-1}\frac{(u-t)!}{(t-1)!0!}\right]$
2	$\frac{t!}{2!u!}\left[\frac{(u-2)!}{0!(t-2)!} - \frac{(u-3)!}{1!(t-3)!} + \cdots + (-1)^{t-2}\frac{(u-t)!}{(t-2)!0!}\right]$
$\cdots$	$\cdots$
t	$\frac{t!}{t!u!}\left[\frac{(u-t)!}{0!0!}\right]$

We now determine the first four moments, ν_1, ν_2, ν_3, and ν_4, of this distribution about the origin $s = 0$. Since, in general,

$$\nu_k = \sum_{s=0}^{t} [s^k \times (\text{Relative frequency of } s)] = \sum_{s=0}^{t} s^k P_{(s)},$$

the tabulation for the computation of any moment is as follows:

Table II: The Computation of the k^{th} Moment of the Distribution of s

s	$s^k P_{(s)}$
0	0
1	$\frac{1^k t!}{1!u!}\left[\frac{(u-1)!}{0!(t-1)!}-\frac{(u-2)!}{1!(t-2)!}+\frac{(u-3)!}{2!(t-3)!}-\cdots+(-1)^{t-1}\frac{(u-t)!}{(t-1)!0!}\right]$
2	$\frac{2^k t!}{2!u!}\left[\frac{(u-2)!}{0!(t-2)!}-\frac{(u-3)!}{1!(t-3)!}+\cdots+(-1)^{t-2}\frac{(u-t)!}{(t-2)!0!}\right]$
3	$\frac{3^k t!}{3!u!}\left[\frac{(u-3)!}{0!(t-3)!}-\cdots+(-1)^{t-3}\frac{(u-2)!}{(t-3)!0!}\right]$
t	$\frac{t^k t!}{t!u!}\left[\frac{(u-t)!}{0!0!}\right]$.

Noting that $\frac{1^k}{1!}=\frac{1^{k-1}}{0!}, \frac{2^k}{2!}=\frac{2^{k-1}}{1!}, \cdots \frac{t^k}{t!}=\frac{t^{k-1}}{(t-1)!}$, and multiplying the terms in brackets by these factors, we develop Table III:

Table III

s	$s^k P_{(s)}$
	1^{st} diagonal ↓ 2^{nd} diagonal ↓ 3^{rd} diagonal ↓
0	0
1	$\frac{t!}{u!}\left[\frac{1^{k-1}(u-1)!}{0!0!(t-1)!}-\frac{1^{k-1}(u-2)!}{0!1!(t-2)!}+\frac{1^{k-1}(u-3)!}{0!2!(t-3)!}\right.$
	t^{th} diagonal ↓ $\left.-\cdots+(-1)^{t-1}\frac{1^{k-1}(u-t)!}{0!(t-1)!0!}\right]$ (t terms)
2	$\frac{t!}{u!}\left[\frac{2^{k-1}(u-2)!}{1!0!(t-2)!}-\frac{2^{k-1}(u-3)!}{1!1!(t-3)!}\right.$
	$\left.+\cdots+(-1)^{t-2}\frac{2^{k-1}(u-t)!}{1!(t-2)!0!}\right]$ (t – 1 terms)
3	$\frac{t!}{u!}\left[\frac{3^{k-1}(u-3)!}{2!0!(t-3)!}-\cdots+(-1)^{t-3}\frac{3^{k-1}(u-t)!}{2!(t-3)!0!}\right]$ (t – 2 terms)
...	...
t	$\frac{t!}{u!}\left[\frac{t^{k-1}(u-t)!}{(t-1)!0!0!}\right]$ (1 term)

Since each series in brackets is one term shorter than the preceding series, the table forms a system of t diagonals. The sum which gives us ν_k may therefore be considered as the sum of these diagonals.

Now, from inspection, it is evident that the general diagonal is of the form

$$s^{\text{th}} \text{ diagonal} = \frac{t!(u-s)!}{u!(t-s)!}\left[\frac{s^{k-1}}{(s-1)!0!} - \frac{(s-1)^{k-1}}{(s-2)!1!} + \cdots + (-1)^{s-1}\frac{1^{k-1}}{0!(s-1)!}\right]$$

$$= \frac{t!(u-s)!}{u!(t-s)!}\left[\frac{1}{(s-1)!}\sum_{r=0}^{s-1}(-1)^r(s-r)^{k-1}\binom{s-1}{r}\right].$$

But it can be shown[4] that

$$\sum_{r=0}^{s-1}(-1)^r(s-r)^{k-1}\binom{s-1}{r} = 0 \qquad \text{when} \qquad k < s.$$

Whence

$$s^{\text{th}} \text{ diagonal} = 0 \qquad \text{when} \qquad k < s.$$

Therefore ν_k is given simply by the sum of the first k diagonals of Table III. Or, in general,

$$\nu_k = \frac{t!(u-1)!}{s!(t-1)!}\left[\frac{1^{k-1}}{0!0!}\right]$$
$$+ \frac{t!(u-2)!}{u!(t-2)!}\left[\frac{2^{k-1}}{1!0!} - \frac{1^{k-1}}{0!1!}\right]$$
$$+ \frac{t!(u-3)!}{u!(t-3)!}\left[\frac{3^{k-1}}{2!0!} - \frac{2^{k-1}}{1!1!} + \frac{1^{k-1}}{0!2!}\right]$$
$$+$$
$$+ \frac{t!(u-k)!}{u!(t-k)!}\left[\frac{k^{k-1}}{(k-1)!0!} - \frac{(k-1)^{k-1}}{(k-2)!1!} + \cdots + (-1)^{k-1}\frac{1^{k-1}}{0!(k-1!)}\right]. \tag{3}$$

To this equation we must, of course, add the condition $k \leqq t$.

[4] E. Netto, *Lehrbuch der Combinatorik*, Leipzig, 1901, 249, Formula 17.

Solving now for the first four moments, we have

$$\nu_1 = \frac{t}{u},$$

$$\nu_2 = \frac{t}{u}\left[1 + \frac{t-1}{u-1}\right],$$

$$\nu_3 = \frac{t}{u}\left[1 + 3\frac{t-1}{u-1} + \frac{(t-1)(t-2)}{(u-1)(u-2)}\right],$$

$$\nu_4 = \frac{t}{u}\left[1 + 7\frac{t-1}{u-1} + 6\frac{(t-1)(t-2)}{(u-1)(u-2)} + \frac{(t-1)(t-2)(t-3)}{(u-1)(u-2)(u-3)}\right].$$

If now we define, for convenience,

$$\overline{u-1}\,,$$

$$t-2$$

$$d = \frac{t-3}{u-3},$$

we have, for the constants of the distribution of s,

$$\text{Mean} = \nu_1 = a.$$

$$\left.\begin{aligned}
\mu_2 &= \nu_2 - \nu_1^2 \\
&= a(1+b) - a^2, \qquad \text{whence } \sigma = \sqrt{a(1+b) - a^2}. \\
\mu_3 &= \nu_3 - 3\nu_1\nu_2 + 2\nu_1^3 \\
&= a(1 + 3b + bc) - 3a^2(1+b) + 2a^3 \\
\mu_4 &= \nu_4 - 4\nu_1\nu_3 + 6\nu_1^2\nu_2 - 3\nu_1^4 \\
&= a(1 + 7b + 6bc + bcd) - 4a^2(1 + 3b + bc) + 6a^3(1+b) - 3a^4.
\end{aligned}\right\} \tag{5}$$

From these constants we can determine the skewness and kurtosis of the distribution of s,

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \quad \text{and} \quad \beta_2 = \frac{\mu_4}{\mu_2^2}. \tag{6}$$

Now it is known that the means of samples of size n drawn from a parent population with constants β_1 and β_2 are distributed in such a way that

$$\beta_{1(\text{means})} = \frac{\beta_1}{n}, \qquad \beta_{2(\text{means})} = 3 + \frac{\beta_2 - 3}{n} \tag{7}$$

Therefore, having determined the beta-constants for the distribution of s, we can determine the beta-constants of the distribution of $\bar{s}$, the mean number of correct matchings resulting from n independent trials.

Now when $t = u \geqq 4$, we have

$$a = b = c = d = 1,$$

and equations (5) give us for the distribution of s

$$\begin{aligned} \text{Mean} &= 1, \\ \mu_2 &= 1, \\ \mu_3 &= 1, \\ \mu_4 &= 4. \end{aligned} \qquad \text{whence} \quad \begin{cases} \beta_1 = 1 \\ \beta_2 = 4 \end{cases}$$

and therefore, for the constants of the distribution of $\bar{s}$, we have, by equations (7),

$$\beta_1 = \frac{1}{n}, \qquad \text{and} \qquad \beta_2 = 3 + \frac{1}{n}$$

which indicates a positively skewed and leptokurtic distribution. The effect of increasing u and holding t constant is to increase the skewness, as shown in the following table for $t = 5$:

t	u	β_1
5	5	$\frac{1}{n}$
5	6	$\frac{1.05}{n}$
5	7	$\frac{1.16}{n}$
5	8	$\frac{1.31}{n}$
5	9	$\frac{1.46}{n}$

The degrees of skewness and kurtosis met with in practical cases of matching with any considerable number of judges (n) are such that a Pearson Type III distribution curve gives a reasonably good fit to the distribution of mean numbers of correct matchings. If, therefore, we have to determine the significance

of any obtained mean number of correct matchings, we may resort to Salvosa's tables[5] of the area under the Type III curve.

As a concrete example of the application of this method let us imagine that 10 judges have arranged 5 character sketches against 8 specimens of handwriting, 5 of which are true apposites of the sketches. Let the total number of correct matchings achieved by this group be 12, whence the mean number per judge is 1.2. We have, then,

$$\bar{s} = 1.2, \qquad n = 10,$$

$$t = 5, \qquad u = 8, \qquad \text{whence} \qquad a = \frac{t}{u} = .625,$$

$$b = \frac{t-1}{u-1} = .571,$$

$$c = \frac{t-2}{u-2} = .500.$$

We now find the mean, standard deviation, and β_1 of the distribution of $\bar{s}$, as follows:

The mean of the distribution of $\bar{s}$ is, by sampling theory, the same as the mean of the distribution of s.

$$\text{Mean} = a = .625.$$

The second moment of the distribution of $\bar{s}$ is, by sampling theory, $\frac{1}{n}$ times the second moment of the distribution of s; whence, by equation (5),

$$\text{Standard deviation} = \sqrt{\frac{1}{10}[a(1+b) - a^2]} = .243.$$

And, by equations (5) and (7),

$$\beta_1 = \frac{1}{10}\frac{[a(1+3b+bc) - 3a^2(1+b) + 2a^3]^2}{[a(1+b) - a^2]^3} = .131.$$

Now the obtained mean number of correct matchings was 1.2, and the next lower number which could have occurred (corresponding to a total of 11 instead of 12 for the group of judges) is 1.1. The lower boundary of the class-interval whose midpoint is $\bar{s} = 1.2$ is therefore 1.15; and it is the area above this boundary under the curve of $\bar{s}$ in which we are interested.

[5] L. R. Salvosa, Tables of Pearson's Type III function, *Ann. Math. Statist.*, 1, 1930, 191–198.

The deviation of this boundary from the mean of $\bar{s}$ is

$$1.15 - .625 = .525\,,$$

and this deviation expressed in terms of the standard deviation gives

$$\frac{.525}{.243} = 2.16\,.$$

Entering Salvosa's table for the deviation 2.16 and skewness $= \sqrt{\beta_1} = .36$, we find by interpolation that so good a performance should be expected by chance only about 23 times in 1000.

MOMENTS ABOUT THE ARITHMETIC MEAN OF A BINOMIAL FREQUENCY DISTRIBUTION

W. J. KIRKHAM, *Oregon State College*

Although the most useful moments of a binomial distribution have been derived as a function of the parameters of the generating binomial for any binomial frequency series, a generalization of notation and procedure is well worth our consideration. The problem attempted in this paper is the calculation of the moments about the mean for the general frequency series of Table I.

TABLE I

The generalized binomial frequency series

x (item)	f (frequency)
0	$N \cdot {}_nC_0 p^0 q^n$
1	$N \cdot {}_nC_1 p^1 q^{n-1}$
2	$N \cdot {}_nC_2 p^2 q^{n-2}$
	$N \cdot {}_nC_n p^n q^0$

In calculating the moments of a set of data about any value, it is often found convenient to use an arbitrary origin, define the moments about this value, and represent the desired moments in terms of those calculated. In the general binomial series, the origin of the x's is found to be the best arbitrary origin. These intermediate moments are

$$\nu_1 = \frac{\sum fx}{N} = M, \text{ arithmetic mean;}$$

$$\nu_2 = \frac{\sum fx^2}{N}; \tag{1}$$

$$\nu_n = \frac{\sum fx^n}{N}$$

where ν_i is the i^{th} moment.

The moments (μ's) about the mean are easily defined as functions of the ν's

from fundamental definitions of the μ's. Denoting the i^{th} moment by μ_i, we have

$$\begin{aligned} \mu_1 &= \frac{\sum f(x - M)}{N} = 0\,, \\ \mu_2 &= \frac{\sum f(x - M)^2}{N} = \nu_2 - \nu_1^2\,, \\ \mu_3 &= \frac{\sum f(x - M)^3}{N} = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3\,, \end{aligned} \tag{2}$$

In general,

$$\mu_n = \nu_n - {}_nC_1\nu_{n-1}\nu_1 + {}_nC_2\nu_{n-2}\nu_1^2 + \cdots + (-1)^{n-1}({}_nC_{n-1} - 1)\nu_1^n\,. \tag{3}$$

Or, if we let $\{\nu\}^n = \nu_n$, we may express the n^{th} moment by a simple notation.

$$\mu_n = \{\mu\}^n = \{\nu\}^n - {}_nC_1\{\nu\}^{n-1}\nu_1 + {}_nC_2\{\nu\}^{n-2}\nu_1^2 + \cdots = (\{\nu\} - \nu_1)^n\,. \tag{4}$$

Solving the equation for $\{\nu\}$,

$$\{\nu\} = \{\mu\} + \nu_1\,.$$

Raising both sides to the n^{th} power and substituting for the "brace" notation,

$$\nu_n = \mu_n + {}_nC_1\mu_{n-1}\nu_1 + {}_nC_2\mu_{n-2}\nu_1^2 + \cdots + \nu_1^n\,.$$

Whence

$$\mu_n = \nu_n - {}_nC_1\mu_{n-1}\nu_1 - {}_nC_2\mu_{n-2}\nu_1^2 - \cdots - \nu_1^n\,, \tag{5}$$

a semi-recursion formula.

The original formula for μ_n contained n moments or variables; and since there are only $(n - 2)$ of the μ's which are of lower order than μ_n, it is necessary to retain ν_n and ν_1 in (5). Since $\mu_1 = 0$, one term in the expansion of μ_n is zero. For instance, when $n = 5$, we have

$$\mu_5 = \nu_5 - 5\mu_4\nu_1 - 10\mu_3\nu_1^2 - 10\mu_2\nu_1^3 - \nu_1^5\,.$$

To calculate μ_k, it is necessary to calculate the ν's from ν_1 to ν_k. For the binomial series, these ν's are

$$\begin{aligned} \nu_1 &= 1npq^{n-1} + \frac{2(n)(n-1)}{1\cdot 2}\, p^2q^{n-2} + \frac{3(n)(n-1)(n-2)}{1\cdot 2\cdot 3}\, p^3q^{n-3} + \cdots \\ &= np\left[q^{n-1} + \frac{(n-1)}{1}\, p^1q^{n-2} + \frac{(n-1)(n-2)}{2!}\, p^2q^{n-3} + \cdots + p^{n-1} \right] \\ &= np(q + p)^{n-1} = np\,, \\ \nu_2 &= np\left[1\cdot q^{n-1} + \frac{2(n-1)}{1!}\, p^1q^{n-2} + 3\,\frac{(n-1)(n-2)}{2!}\, p^2q^{n-3} + \cdots + np^{n-1} \right], \end{aligned}$$

$$\nu_3 = np\left[1^2\cdot q^{n-1} + 2^2\frac{(n-1)}{1!}p^1q^{n-2} + 3^2\frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \cdots + n^2p^{n-1}\right],$$

$$\nu_k = np\left[1^{k-1}q^{n-1} + 2^{k-1}\frac{(n-1)}{1!}p^1q^{n-}\right.$$

$$\left. + 3^{k-1}\frac{(n-1)(n-2)}{2!}p^2q^{n-3} + \cdots + n^{k-1}p^{n-1}\right].$$

In the simplified form of ν_k, the [ ] is the $(k - 1)^{\text{th}}$ moment about -1 of the binomial series generated by the binomial $(q + p)^{n-1}$. Denoting this [ ] by $\nu'_{k-1}(n - 1)$, the ν's can be expressed by the formula

$$\nu_k = np\nu'_{k-1}(n - 1), \tag{6}$$

where ν' is a function of $(n - 1)$ and $(k - 1)$ while ν_k was a function of n and k.

Let us see how a ν' in ν_k can be defined in terms of the ν's of lower order than k. In finding this relationship, a consideration of the two series of Table II will be helpful.

TABLE II

x'	f	x	f
1	$N_{n-1}C_0p^0q^{n-1}$	0	$N_{n-1}C_0p^0q^{n-1}$
2	$N_{n-1}C_1p^1q^{n-2}$	1	$N_{n-1}C_1p^1q^{n-2}$
.	....	.	
.	....	.	
n	$N_{n-1}C_{n-1}p^nq^0$	$n -$	$N_{n-1}C_{n-1}p^{n-1}q^0$

The [ ] in ν_k for Table I is equal to the $(k - 1)^{\text{th}}$ moment of x' about $x' = 0$. Or

$$\nu_{k-1}, \text{ Table II}, x', = \nu_{k-1}, \text{ Table I}, = \nu'_{k-1}(n - 1).$$

Also ν_{k-1} for x, Table II, is ν_{k-1} for the series generated by $(q - p)^{n-1}$.

The desired relationship between the ν's for the two series of Table II can be found by making use of the equations expressing the equality of the μ's for x and x'. Dropping the variable which shows the number of items, the same for the two series of Table II, in the notation, we have

$$\mu_2 = \mu'_2 = \nu_2 - \nu_1^2 = \nu'_2 - \nu_2'^2, \qquad \nu'_2 = \nu_2 - 2\nu_1(\nu_1 - \nu'_1) + (\nu_1 - \nu'_1)^2;$$

$$\mu_3 = \mu'_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 = \nu'_3 - 3\nu'_2\nu'_1 + 2\nu_1'^3,$$

$$\nu'_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 + 3\nu'_2\nu'_1 - 2\nu_1'^3.$$

Substituting the value of ν'_2 in the right member of ν'_3,

$$\nu'_3 = \nu_3 - 3\nu_2(\nu_1 - \nu'_1) + 3\nu_1(\nu_1 - \nu'_1)^2 - (\nu_1 - \nu'_1)^3.$$

In general,

$$\nu_k' = \nu_k - {}_kC_1\nu_{k-1}(\nu_1 - \nu_1') + {}_kC_2\nu_{k-2}(\nu_1 - \nu_1')^2 + \cdots + (-1)^k(\nu_1 - \nu_1')^k. \tag{7}$$

The formula just derived may be used to define the moments about any origin in terms of those about the original zero of the x's. For our immediate use, the formula simplifies since $\nu_1' = \nu_1 + \nu_0 = \nu_1 + 1$. Then

$$\nu_k' = \nu_k + {}_kC_1\nu_{k-1} + {}_kC_2\nu_{k-2} + {}_kC_3\nu_{k-3} + \cdots \tag{8}$$

By simple analysis we found the value of ν_1 to be np. By the method of continuation, we are able to extend the list of ν's to any number. ν' from (8) is used in (6) with n replaced by $(n - 1)$ in the ν's.

$\nu_0 = 1.$

$\nu_1 = np.$

$$\begin{aligned}
\nu_2 &= np\nu_1'(n-1) = np[\nu_1(n-1) + \nu_0(n-1)] \\
&= np[(n-1)p + 1] = n(n-1)p^2 + np. \\
\nu_3 &= np\nu_2'(n-1) = np[\nu_2(n-1) + 2\nu_1(n-1) + \nu_0(n-1)] \\
&= n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np. \\
\nu_4 &= np\nu_3'(n-1) = np[\nu_3(n-1) + 3\nu_2(n-1) + 3\nu_1(n-1) + \nu_0(n-1)] \\
&= np\{[(n-1)(n-2)(n-3)p^3 + 3(n-1)(n-2)p^2 + (n-1)p] \\
&\qquad + 3[(n-1)(n-2)p^2 + (n-1)p] + 3[(n-1)p] + 1\}. \\
&= n(n-1)(n-2)(n-3)p^4 + 6(n)(n-1)(n-2)p^3 + 7(n)(n-1)p^2 + np.
\end{aligned}$$

. . . .

If the order of the terms in the expansion is reversed, ν_n is an ascending power series in p. The pure numerical coefficients in some of these ν's are

$\nu_1 = (1)$

$\nu_2 = (1, 1)$

$\nu_3 = (1, 3, 1)$

$\nu_4 = (1, 7, 6, 1)$

$\nu_5 = (1, 15, 25, 10, 1)$

$\nu_6 = (1, 31, 90, 65, 15, 1)$

$\nu_7 = (1, 63, 301, 350, 140, 21, 1)$

$\nu_8 = (1, 127, 966, 1701, 1050, 266, 28, 1).$

In general,

$$\nu_{n+1} = \left(1, \sum_1^n {}_nC_i, \sum_2^n \left({}_nC_i \sum_1^{i-1} {}_{i-1}C_j\right), \right.$$
$$\left. \sum_3^n \left({}_nC_i \sum_2^{i-1} \left({}_{i-1}C_j \sum_1^{j-1} {}_{j-1}C_k\right)\right), \cdots\right). \tag{9}$$

Using the foregoing ν's, and the semi-recursion formula, we are able to determine the μ's.

$$\begin{aligned}
\mu_2 &= \nu_2 - \nu_1^2 \\
&= [np + (n)(n-1)p^2] - (np)^2 \\
&= np(1-p) \\
&= npq.
\end{aligned}$$

$$\begin{aligned}
\mu_3 &= \nu_3 - 3\nu_1\mu_2 - \nu_1^3 \\
&= [np + 3n(n-1)p^2 + (n)(n-1)(n-2)p^3] - 3(np)[np(1-p)] - [np]^3. \\
&= np + (-3n)p^2 + (2n)p^3 = np(1 - 3p + 2p^2) \\
&= np(1-p)(1-2p) \\
&= npq(q-p).
\end{aligned}$$

$$\begin{aligned}
\mu_4 &= [np + 7(n)(n-1)p^2 + 6(n)(n-1)(n-2)p^3 + (n)(n-1)(n-2) \\
&\qquad (n-3)p^4] - 4(np)(np)(1 - 3p + 2p^2) - 6(np)^2(np)(1-p) - (np)^4 \\
&= np + (-7n + 3n^2)p^2 + (12n - 6n^2)p^3 + (-6n + 3n^2)p^4 \\
&= np(1 - 7p + 12p^2 - 6p^3) + 3n^2p^2(1 - 2p + p^2) \\
&= npq - 6np^2q^2 + 3n^2p^2q^2.
\end{aligned}$$

$$\begin{aligned}
\mu_5 &= np(1 - 15p + 50p^2 - 60p^3 + 24p^4) + 10n^2p^2(1 - 4p + 5p^2 - 2p^3) \\
&= (q-p)(npq - 12np^2q^2 + 10n^2p^2q^2).
\end{aligned}$$

$$\begin{aligned}
\mu_6 &= np(1 - 31p + 180p^2 - 390p^3 + 360p^4 - 120p^5) + 5n^2p^2(5 - 36p \\
&\qquad + 83p^2 - 78p^3 + 26p^4) + 15n^3p^3(1 - 3p + 3p^2 - p^3) \\
&= npq - 30np^2q^2(q-p)^2 + 25n^2p^2q^2 - 130n^2p^3q^3 + 15n^3p^3q^3.
\end{aligned}$$

$$\begin{aligned}
\mu_7 &= np(1 - 63p + 602p^2 - 2100p^3 + 3360p^4 - 2520p^5 + 720p^6) \\
&\qquad + n^2p^2(56 - 686p + 2590p^2 - 4270p^3 + 3234p^4 - 924p^5) + n^3p^3(105 \\
&\qquad - 525p + 945p^2 - 735p^3 + 210p^4)
\end{aligned}$$

$$= (q - p)(npq - 60np^2q^2 + 360np^3q^3 + 56n^2p^2q^2 - 462n^2p^3q^3 + 105n^3p^3q^3).$$

$$\begin{aligned}
\mu_8 = {} & np(1 - 127p + 1932p^2 - 10206p^3 + 25200p^4 - 31920p^5 + 20160p^6 \\
& - 5040p^7) + n^2p^2(119 - 2394p + 13895p^2 - 35700p^3 + 46004p^4 \\
& - 29232p^5 + 7308p^6) + n^3p^3(490 - 3850p + 10990p^2 - 14770p^3 \\
& + 9520p^4 - 2380p^5) + n^4p^4(105 - 420p + 630p^2 - 420p^3 + 105p^4) \\
= {} & npq(1 - 42pq(3 - 10pq(1 - 3pq))) + 7n^2p^2q^2(17 - 4pq(77 - 261pq)) \\
& + 70n^3p^3q^3(7 - 34pq) + 105n^4p^4q^4.
\end{aligned}$$

ON CERTAIN DISTRIBUTION FUNCTIONS WHEN THE LAW OF THE UNIVERSE IS POISSON'S FIRST LAW OF ERROR[1]

By Frank M. Weida

Introduction. The median, which is that value of a permuted variable which has as many observed values on one side of it as on the other, appears to be the natural competitor of the arithmetic mean when we are interested in the probable or most probable value of an unknown quantity. It is well known[2] that the law of probability, namely, Poisson's first law of error, which results from the assumption that the median is the most probable value of the unknown quantity is given by

$$f(x) = \frac{k}{\sigma} e^{-\frac{|x|}{\sigma}}. \tag{1}$$

Little is known about the form of the distribution functions of the more important statistics when the law of the "Universe" is Poisson's first law of error. It, therefore, appears to be of interest and importance to enlarge our present knowledge of distribution functions by finding certain new ones when the variable or variables are defined by (1).

In this paper we present the following results: (1) We have obtained an explicit expression for the distribution of means of samples of n; (2) we have obtained an explicit expression for the distribution of differences; (3) we have obtained an explicit expression for the distribution of quotients; (4) we have obtained an explicit expression for the distribution of standard deviations for samples of n; (5) we have obtained an explicit expression for the distribution of geometric means for samples of n; (6) we have obtained an explicit expression for the distribution of harmonic means for samples of n.

In our analysis, we have made use of the theory of characteristic functions in the sense of Levy.[3] This theory has been extended to more than one dimension by V. Romanovsky[4] and by E. K. Haviland.[5] S. Kullback,[6] in his thesis, has made further extensions and has applied them successfully to the distribution problem in statistics.

[1] Presented to the American Mathematical Society, February 23, 1935.

[2] Brunt, David: "The Combination of Observations," 1923, p. 27.

[3] Levy, P.: "Calcul des Probabilités;" pp. 153–191.

[4] Romanovsky, V.: "Sur un théorème limite du calcul des probabilités," Recueil mathématique de la Société mathématique de Moscow, Vol. 36, 1926, pp. 36–64.

[5] Haviland, E. K.: "On the inversion formula for Fourier-Stieltjes transforms in more than one dimension," American Journal of Mathematics, Vol. 57, 1935, pp. 94–101.

[6] Kullback, S.: "An application of characteristic functions to the distribution problem of statistics," Annals of Mathematical Statistics, Vol. V, No. 4, pp. 263–307.

The explicit expression for the distribution of arithmetic means of samples of n is not new. This law of distribution has previously been obtained otherwise by F. Hausdorff[7] and by A. T. Craig.[8] It is inserted here to show the superiority and greater power of our method when compared with previous methods and for the completeness of our discussion. The other results offered in this paper, as far as the writer knows, are new.

1. The distribution of arithmetic means. Let us consider

$$f(x) = \frac{k}{\sigma} e^{-\frac{|x|}{\sigma}}, \qquad (-\alpha < x < \alpha). \tag{2}$$

If we assume that $x_1, x_2, \cdots, x_n$ are independently distributed and that each $x_i (i = 1, 2, \cdots, n)$ is distributed according to the same distribution law, namely, Poisson's first law of error, then it is fairly easy to see that the characteristic function for the law of distribution of means of samples of n is given by

$$\phi(t) = \left\{ \int_{-\alpha}^{\alpha} \frac{k}{\sigma} e^{it|x| - \frac{|x|}{\sigma}} dx \right\}^n .$$

If $u = \Sigma_i x_i$ $(i = 1, 2, \cdots, n)$, then it follows that the distribution function of u, namely, $P(u)$, is given by

$$P(u) = \frac{1}{2\pi} \int_{\alpha}^{\alpha} e^{-itu} \left\{ \frac{2k}{\sigma} \int_0^{\alpha} e^{itx - \frac{x}{\sigma}} dx \right\}^n dt, \tag{4}$$

which, upon simplification becomes

$$P(u) = \frac{2^{n-1} k^n}{\pi \sigma^n} \int_{-\alpha}^{\alpha} \frac{e^{-itu} \, dt}{(1 - \sigma i t)^n} . \tag{5}$$

It is readily seen that the poles of the integrand are of the n^{th} order and are those of $(1 - \sigma i t)^n$. It follows by the well known Residue Theorem of Cauchy[9] that

$$P(u) = \frac{2^{n-1} k^n}{\pi \sigma^n} \cdot 2\pi i \cdot \frac{(-1)^{n-1}}{(n-1)!} \cdot \frac{1}{i^n} \cdot \frac{d^{n-1}}{dt^{n-1}} \left\{ \frac{e^{-itu}}{(1 + \sigma i t)^n} \right\}_{t = \frac{1}{\sigma i}} . \tag{6}$$

If now, we replace u by $n|\bar{x}|$, we will obtain the desired law of the distribution of arithmetic means of samples of n which is

$$P(|\bar{x}|) = \frac{2^n k^n (-1)^{n-1} n}{\sigma^n i^{n-1} (n-1)!} \cdot \frac{d^{n-1}}{dt^{n-1}} \left\{ \frac{e^{-itn|x|}}{(1 + \sigma i t)^n} \right\}_{t = \frac{1}{\sigma i}} \tag{7}$$

defined for all values of x on the range $(-\alpha < x < \alpha)$.

[7] Hausdorff, F.: Beitrage zur Wahrscheinlichkeitsrechnung Königlich Sachsischen Gesellschaft der Wissenshaften zu Leipzig. Berichted über die Verhandlungen Math - Phys. Classe, Vol. 53, 1901, pp. 152–178.

[8] Craig, A. T.: "On the distribution of certain statistics," American Journal of Mathematics, Vol. 54, 1932, pp. 353–366.

[9] Macrobert, T. M.: "Functions of a Complex Variable," 1933, pp. 57, 295.

A. T. Craig[8] has given the distribution laws of arithmetic means of samples of size 2, 3, and 4. These results as well as the results for any n are readily obtained from (7).

2. **The distribution of differences.** Let us assume that the laws of distribution of x and y are independent and that they are given respectively by

$$f(x) = \frac{k_1}{\sigma_1} e^{-\frac{|x|}{\sigma_1}}; \qquad f(y) = \frac{k_2}{\sigma_2} e^{-\frac{|y|}{\sigma_2}}; \qquad (-\alpha < x < \alpha), \qquad (-\alpha < y < \alpha).$$

In this case, the characteristic function of the law of distribution of differences $(x - y)$ is given by

$$\phi(t) = \frac{k_1}{\sigma_1} \int_{-\alpha}^{\alpha} e^{it|x| - \frac{|x|}{\sigma_1}} dx \cdot \frac{k_2}{\sigma_2} \int_{-\alpha}^{\alpha} e^{-it|y| - \frac{|y|}{\sigma_2}} dy. \tag{8}$$

Performing the operations indicated in (8) and simplifying, we find that

$$\phi(t) = \frac{4k_1k_2}{\sigma_1\sigma_2} \cdot \frac{1}{(1 - \sigma_1 it)} \cdot \frac{1}{(1 + \sigma_2 it)}. \tag{9}$$

It is fairly easy to see that the distribution law of u is given by

$$P(u) = \frac{4k_1k_2}{2\pi\sigma_1\sigma_2} \int_{-\alpha}^{\alpha} \frac{e^{-itu}\, dt}{(1 - \sigma_1 it)(1 + \sigma_2 it)}. \tag{10}$$

Now, let $\{(1/\sigma_1) - it\} = v/u$, then (10) becomes

$$P(u) = \frac{2k_1k_2 e^{-\frac{u}{\sigma_1}}}{\pi i\sigma_1\sigma_2(\sigma_1 + \sigma_2)} \int_{-\frac{u}{\sigma_1} - i\alpha}^{-\frac{u}{\sigma_1} + i\alpha} \frac{e^{-v}\, dv}{(-v)\left\{1 + \frac{}{\left(\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2} u\right)}\right\}} \tag{11}$$

The integral in (11) is convergent because

$$\operatorname{Lim} v^m \frac{e^{-v}\, dv}{(-v)\left\{1 + \left(\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2} u\right)\right.} = 0.$$

Hence, we find that

$$P(u) = -\frac{2k_1k_2 e^{-\frac{u}{\sigma_1}}}{\pi i\sigma_1\sigma_2(\sigma_1 + \sigma_2)} \int_{\alpha}^{(0+)} \frac{e^{-v}\, dv}{(-v)\left\{1 + \frac{v}{\left(\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2} u\right)}\right\}} \tag{12}$$

which upon simplification becomes

$$P(u) = \frac{4k_1k_2}{\sigma_1\sigma_2(\sigma_1 + \sigma_2)} W_{0,\frac{1}{2}}\left\{\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2} u\right\}, \tag{13}$$

where $W_{0,\frac{1}{2}}\left\{\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2} u\right\}$ is the *confluent hypergeometric function.*[10]

It is well known that

$$W_{k,m}(z) = \frac{e^{-\frac{1}{2}z} z^k}{\Gamma(\frac{1}{2} - k + m)} \int_0^{\alpha} t^{-k-\frac{1}{2}+m}\left(1 + \frac{t}{z}\right)^{k-\frac{1}{2}+m} e^{-t}\, dt$$

for all values of k and m and for all values of z except negative real values. Clearly,

$$W_{0,\frac{1}{2}}\left\{\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2} u\right\} = \frac{e^{-\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\frac{1}{2}u}}{\Gamma(1)} \int_0^{\alpha} e^{-t}\, dt$$

which, upon simplification becomes

$$W_{0,\frac{1}{2}}\left\{\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2} u\right\} = e^{-\frac{\sigma_1+\sigma_2}{2\sigma_1\sigma_2}u}. \tag{14}$$

Hence, we now find that

$$P(u) = \frac{4k_1k_2}{\sigma_1\sigma_2(\sigma_1 + \sigma_2)} e^{-\frac{\sigma_1+\sigma_2}{2\sigma_1\sigma_2}u}. \tag{15}$$

If now, we replace u by $|x| - |y|$, we will obtain the desired law of distribution of differences which is

$$P(|x| - |y|) = \frac{4k_1k_2}{\sigma_1\sigma_2(\sigma_1 + \sigma_2)} e^{-\frac{\sigma_1+\sigma_2}{2\sigma_1\sigma_2}\{|x| - |y|\}}. \tag{16}$$

3. **The distribution of ratios.** We assume that the laws of distribution of x and y are independent and that they are given respectively by

$$f(x) = \frac{k_1}{\sigma_1} e^{-\frac{|x|}{\sigma_1}}; \quad f(y) = \frac{k_2}{\sigma_2} e^{-\frac{|y|}{\sigma_2}}; \quad (-\alpha < x < \alpha), \quad (-\alpha < y < \alpha).$$

[10] Whittaker, E. T. and Watson, G. N.: "A course in modern Analysis," 1915, pp. 333-334.

Let $u = \log|x| - \log|y|$. The characteristic function of the law of distribution of quotients is then given by

$$\begin{aligned}\phi(t) &= \frac{k_1}{\sigma_1}\int_{-\alpha}^{\alpha} e^{-\frac{|x|}{\sigma_1}}(|x|)^{it}\,dx\cdot\frac{k_2}{\sigma_2}\int_{-\alpha}^{\alpha} e^{-\frac{|y|}{\sigma_2}}(|y|)^{-it}\,dy \\ &= \frac{4k_1k_2}{\sigma_1\sigma_2}\int_0^{\alpha} e^{-\frac{x}{\sigma_1}}x^{it}\,dx\int_0^{\alpha} e^{-\frac{y}{\sigma_2}}y^{-it}\,dy.\end{aligned} \tag{17}$$

Now, let $s = x/\sigma_1$ and $w = y/\sigma_2$, then clearly

$$\phi(t) = 4k_1k_2\sigma_1^{it}\sigma_2^{-it}\int_0^{\alpha} e^{-s}s^{it}\,ds\int_0^{\alpha} e^{-w}w^{-it}\,dw,$$

whence

$$\phi(t) = 4k_1k_2\sigma_1^{it}\sigma_2^{-it}\,\Gamma(it+1)\,\Gamma(1-it). \tag{18}$$

It follows that the distribution law of u is given by

$$P(u) = \frac{4k_1k_2}{2\pi}\int_{-\alpha}^{\alpha} e^{-itu+i\log\sigma_1 t-i\log\sigma_2 t}\,\Gamma(it+1)\,\Gamma(1-it)\,dt$$

which upon simplification, becomes

$$P(u) = \frac{2k_1k_2}{\pi}\int_{-\alpha}^{\alpha} e^{-i(u-\log\sigma_1+\log\sigma_2)t}\,\Gamma(it+1)\,\Gamma(1-it)\,dt. \tag{19}$$

Now, let $(1 - it) = -v$, then (19) becomes

$$P(u) = \frac{4k_1k_2}{2\pi i}\int_{-1-i\alpha}^{-1+i\alpha} e^{-v\{u-\log\sigma_1+\log\sigma_2\}-\{u-\log\sigma_1+\log\sigma_2\}}\,\Gamma(2+v)\,\Gamma(-v)\,dv. \tag{20}$$

Since it can be shown that[11]

$$(1/2\pi i)\int_{-1-i\alpha}^{-1+i\alpha} e^{-vu}\Gamma(2+v)\,\Gamma(-v)\,dv = \Gamma(2)\{1+(1/e^u)\}^{-2},$$

we find that (20) becomes

$$P(u) = \frac{4k_1k_2e^{-u}\sigma_1}{\sigma_2}\,\Gamma(2)\left\{1+\frac{\sigma_1}{\sigma_2e^u}\right\}^{-2}. \tag{21}$$

Now, put $e^u = |x|/|y| = R$, whence from (21) we will obtain the desired law of distribution of quotients which is

$$P(R) = \frac{4k_1k_2\sigma_1\Gamma(2)}{\sigma_2R}\left\{1+\frac{\sigma_1}{\sigma_2R}\right\}^{-2}. \tag{22}$$

[11] Macrobert, T. M., "Functions of a Complex Variable," 1933, pp. 114, 139, 151. Whittaker, E. T. and Watson, G. N., "A course in modern Analysis," 1915, pp. 283.

4. **The distribution of variances and standard deviations.** If we assume that the variance and standard deviation are calculated about a sample mean and if we let $u = \sum_{i=1}^{n-1} x_i^2$, and if the x_i are independently distributed and each x_i is distributed according to the same distribution law, namely, Poisson's first law of error, then it is clear that the characteristic function for the law of distribution of variances of samples of n is

$$\phi(t) = \left\{\frac{k}{\sigma}\int_{-\infty}^{\infty} e^{-\frac{|x|}{\sigma}+itx^2}\,dx\right\}^{n-1} = \left\{\frac{2k}{\sigma}\int_0^{\infty} e^{itx^2-\frac{x}{\sigma}}\,dx\right\}^{n-1}. \tag{23}$$

Let I represent the integral in the right-hand member of (23). We obtain that $(dI/d\sigma) = I/\sigma^2$, whence $I = Ce^{-\frac{1}{\sigma}}$. Making use of the conditions:

$$\sigma \to \infty, \qquad\qquad I \to \int_0^{\infty} e^{itx^2}\,dx = e^{\frac{1}{4}\pi i}\frac{\sqrt{\pi}}{\sqrt{t}},$$

$\sigma \to \infty$, $Ce^{-\frac{1}{\sigma}} \to C$, whence we find that

$$\int_0^{\infty} e^{itx^2-\frac{x}{\sigma}}\,dx = e^{\frac{1}{4}\pi i}\frac{\sqrt{\pi}}{\sqrt{t}}\,e^{-\frac{1}{\sigma}}.$$

Clearly, it follows that

$$\phi(t) = \frac{2^{n-1}k^{n-1}e^{\frac{(n-1)\pi i}{4}}}{\sigma^{n-1}t^{\frac{n-1}{2}}}\,\pi^{\frac{n-1}{2}}\,e^{-\frac{n-1}{\sigma}}. \tag{24}$$

We now find that the distribution law of u is given by

$$P(u) = \frac{2^{n-1}k^{n-1}e^{\frac{(n-1)\pi i}{4}}\pi^{\frac{n-1}{2}}e^{-\frac{n-1}{\sigma}}}{2\pi\sigma^{n-1}}\int_{-\infty}^{\infty}\frac{e^{-iut}}{t^{\frac{n-1}{2}}}\,dt. \tag{25}$$

Evaluating the integral in (25) with a suitably chosen contour,[12] we find that

$$P(u) = \frac{2^{n-1}k^{n-1}\pi^{\frac{n-1}{2}}e^{-\frac{n-1}{\sigma}}2\pi}{2\pi\sigma^{n-1}\,\Gamma\left(\frac{n-1}{2}\right)}\,u^{\frac{n-3}{2}}e^{-u}. \tag{26}$$

Now, let $u = \sum_{i=1}^{n} x_i^2 = ns^2$, whence from (26) we will obtain the desired law of distribution of variances which is

$$P(s^2) = \frac{2^{n-1}k^{n-1}\pi^{\frac{n-1}{2}}e^{-\frac{n-1}{\sigma}}}{\sigma^{n-1}\,\Gamma\left(\frac{n-1}{2}\right)}\,n^{\frac{n-3}{2}}(s^2)^{\frac{n-3}{2}}e^{-s^2}. \tag{27}$$

[12] Macrobert, T. M., "Functions of a Complex Variable," 1933, p. 67.

The law of distributions of standard deviations can be obtained at once from (27) since $d(s^2) = 2s\,ds$.

We shall now give the specific laws of distribution of variances for samples of size 1, 2, 3, 4, and 5 when the law of the "Universe" is Poisson's first law of error. From (27),

For $n = 1$,

$$P(s^2) = 0, \qquad (0 < s^2 < \infty). \tag{28}$$

For $n = 2$,

$$P(s^2) = \frac{2^{\frac{1}{2}} k e^{-\frac{1}{\sigma}} e^{-s^2}}{\sigma s}, \qquad (0 < s^2 < \infty). \tag{29}$$

For $n = 3$,

$$P(s^2) = \frac{4k^2 \pi e^{-\frac{2}{\sigma}} e^{-s^2}}{\sigma^2}, \qquad (0 < s^2 < \infty). \tag{30}$$

For $n = 4$,

$$P(s^2) = \frac{32k^3 \pi e^{-\frac{}{\sigma}} s e^{-s^2}}{\sigma^3}, \qquad (0 < s^2 < \infty). \tag{31}$$

For $n = 5$,

$$P(s^2) = \frac{80k^4 \pi^2 e^{-\frac{4}{\sigma}} s^2 e^{-s^2}}{\sigma^4}, \qquad (0 < s^2 < \infty). \tag{32}$$

5. **The distribution of geometric means.** As before, we assume that the x_i are independently distributed and each x_i is distributed according to the same distribution law, namely, Poisson's first law of error. Then, clearly, the characteristic function for the law of distribution of geometric means of samples of n is

$$\phi(t) = \left\{ \int_{-\alpha}^{\alpha} \frac{k}{\sigma} e^{-\frac{|x|}{\sigma}} |x|^{it}\, dx \right\}^n = \left\{ \frac{2k}{\sigma} \int_0^{\alpha} e^{-\frac{x}{\sigma}} x^{it}\, dx \right\}^n. \tag{33}$$

Now, put $s = x/\sigma$, then (33) becomes

$$\phi(t) = \left\{ 2k\sigma^{it} \int_0^{\alpha} e^{-s} s^{it}\, ds \right\}^n = 2^n k^n \sigma^{nit} \{\Gamma(it + 1)\}^n. \tag{34}$$

It follows at once that the distribution law of u is

$$P(u) = \frac{2^n k^n}{2\pi} \int_{-\alpha}^{\alpha} e^{-i(u + n \log \sigma)t} \{\Gamma(it + 1)\}^n\, dt. \tag{35}$$

Now, let $it + 1 = -v$, then (35) becomes

$$P(u) = \frac{-2^n k^n}{2\pi i} e^{u+n\log\sigma} \int_{-1-i\alpha}^{-1+i\alpha} e^{v(u+n\log\sigma)} \{\Gamma(-v)\}^n \, dv. \tag{36}$$

It is well known that (10)

$$\{\Gamma(-v)\}^n = \frac{(-1)^n \pi^n}{\sin^n \pi v \{\Gamma(v+1)\}^n}. \tag{37}$$

Using (37) in (36), we readily find that

$$P(u) = \frac{-2^n k^n}{2\pi i} e^{u+n\log\sigma} \int_{-1-i\alpha}^{-1+i\alpha} \frac{e^{v(u+n\log\sigma)}(-1)^n \pi^n}{\{\Gamma(v+1)\}^n \sin^n \pi v} \, dv. \tag{38}$$

It is fairly easy to see that the poles of the integrand in (38) are the poles of $\{\Gamma(-v)\}^n$ and that these poles are of the n^{th} order. Applying the well known Residue Theorem of Cauchy (8), we find that

$$P(u) = 2^n k^n e^{u+n\log\sigma} \sum_{a=0}^{\alpha} \frac{(-1)^{n+na+1}}{(n-1)!} \left\{ \frac{d^{n-1}}{dv^{n-1}} \left[\frac{e^{v(u+n\log\sigma)}}{\{\Gamma(v+1)\}^n} \right] \right\}_{v=a}. \tag{39}$$

Now, since $u = \log|x_1| + \log|x_2| + \cdots + \log|x_n|$, then clearly, the distribution law of the geometric mean, G, is obtained from the law of distribution for u by means of the transformation

$$u = \log (G)^n.$$

Hence, from (39), we find the desired law of distribution of geometric means of samples of n which is

$$P\{G\} = \frac{2^n k^n G^n \sigma^n}{\Gamma(n)} \sum_{a=0}^{\alpha} (-1)^{n+na+1} \left\{ \frac{d^{n-1}}{dv^{n-1}} \left[\frac{G^{nv} \sigma^{nv}}{\{\Gamma(v+1)\}^n} \right] \right\}_{v=a}. \tag{40}$$

6. **The distribution of harmonic means.** Let us assume that $f(x)$ is the law of distribution for x. It is well known[13] that the law of distribution of $x' = 1/x$ is given by

$$F(x') = (1/x'^2)\, f(1/x')$$

if $1/x$ is continuous on the range of definition of $f(x)$. Now, in case $f(x)$ is Poisson's first law of error, we find that

$$F(x') = F(1/x) = \frac{k}{\sigma} x^2 e^{-\frac{|x|}{\sigma}}; \qquad (-\alpha \leqq x < 0), \quad (0 < x \leqq \alpha). \tag{41}$$

[13] Dodd, E. L., "The frequency law of a function of one variable," Bulletin of the American Mathematical Society, Vol. 31, 1925, p. 28; "The frequency law of a function of variables with given frequency laws," Annals of Mathematics, Second Series, Vol. 27, 1925–26, p. 18.

We assume that the x_i' are independently distributed and each x_i' is dis-distributed according to the same law of distribution, whence we find that the characteristic function for the law of distribution of harmonic means of samples of n is

$$\phi(t) = \left\{ \int_0^\alpha \frac{k}{\sigma} e^{it|x| - \frac{|x|}{\sigma}} x^2 \, dx \right\}^n, \tag{42}$$

from which, after simplification, we find that

$$\phi(t) = \frac{k^n 2^n \sigma^{2n}}{(1 - \sigma it)^{3n}}. \tag{43}$$

We now find that the law of distribution for u is

$$P(u) = \frac{2^n k^n \sigma^{2n}}{2\pi} \int_{-\alpha}^{\alpha} \frac{e^{-itu}}{(1 - \sigma it)^{3n}} \, dt,$$

which, after evaluation and simplification, becomes

$$P(u) = \frac{2^n k^n}{\sigma^n \Gamma(3n)} u^{3n-1} e^{-\frac{u}{\sigma}}. \tag{44}$$

Recalling that in this case, $u = 1/|x_1| + 1/|x_2| + \cdots + 1/|x_n|$, we make the transformation $u = n/H$, where H is the harmonic mean; whence, from (44), we find that the desired law of distribution of harmonic means of samples of n is given by

$$P(H) = \frac{2^n k^n n^{3n-1}}{\sigma^n \Gamma(3n)} \cdot H^{1-3n} e^{-\frac{n}{\sigma H}}. \tag{45}$$

7. **Conclusions.** We have shown that the same analysis is applicable to find the explicit expression for all the distribution laws we have discussed in this paper.

THE GEORGE WASHINGTON UNIVERSITY,
WASHINGTON, D. C.

ON THE PROBLEM OF CONFIDENCE INTERVALS

By J. Neyman

When discussing my paper read before the Royal Statistical Society on 19th June, 1934, Professor Fisher said that the extension of his work concerning the fiducial argument to the case of discontinuous distributions, as presented in my paper, has been reached at a great expense: that instead of exact probability statements we get only statements in the form of inequalities.

This remark raises the question whether the disadvantage of the solution which he mentioned (the inequalities instead of equalities) results from the unsatisfactory method of approach, or whether it is connected with the nature of the problem itself.

I think that the problem is of considerable general interest. For instance it may be asked whether the confidence intervals for the binomial distribution recently published by E. S. Pearson and C. J. Clopper,[1] which correspond to the probability statements in inequalities, could be bettered.

The purpose of the present note is to show, (1) that in some exceptional cases the exact probability solution of the problem exists and that then it may easily be found by the method described in Note I of my paper;[2] (2) that in the general case of discontinuous distribution exact probability statements in the problem of confidence intervals are impossible.

In particular it will be seen that exact probability statements are impossible in the case of the binomial distribution and so that the system of confidence intervals published by Clopper and Pearson could not be bettered.

In order to avoid any possible misunderstanding I shall start by restating the problem.

We shall consider a random discontinuous variate x, capable of having one or another of a finite, or at most denumerable set of values

$$x_1, x_2, \cdots x_n, \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

We shall assume that the frequency function, say $p(x \mid \theta)$, of x depends upon one parameter θ, the value of which is unknown. The problem of confidence intervals consists in ascribing to every possible value of x e.g. to x_n, $(n = 1, 2, \cdots)$ a "confidence interval," say $\theta_1(n)$ to $\theta_2(n)$ such that the probability, P, of our being correct in stating

$$\theta_1(n) \leqq \theta \leqq \theta_2(n) \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

whenever we observe $x = x_n$ $(n = 1, 2, \cdots)$, is either:

[1] E. S. Pearson and C. J. Clopper: The Use of Confidence or Fiducial Limits in the Case of the Binomial. Biometrika Vol. XXVI, pp. 404–413

[2] J. R. S. S. Vol. 97, p. 589.

(a) equal to a given value $\alpha < 1$ chosen in advance, or

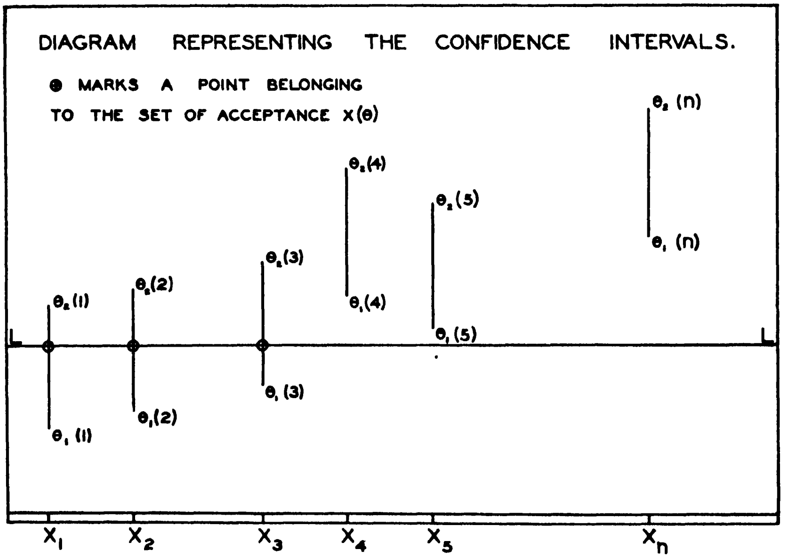
(b) *at least* equal to this value α.

I proposed to call this chosen value α the confidence coefficient.

In the earlier paper I showed that the solution of the problem in its form (b) is always possible and easy to find. If the variate x is continuous, then the solution of the problem (a) is equally easy. At present we shall consider whether and under what conditions the solution (a) is possible when the variate x is discontinuous.

Suppose that the variate x is discontinuous as described above, and that the solution of the problem in its form (a) exists and is given by the system of confidence intervals $(\theta_1(x_n), \theta_2(x_n))$ for $n = 1, 2, \cdots$.

The position is illustrated in the diagram below. On the axis of abscissae the possible values of the variate x are marked. The axis of ordinates is the axis of θ. The confidence intervals are marked on verticals passing through corresponding values of x.



According to our hypothesis the intervals $(\theta_1(x_n), \theta_2(x_n))$ are so chosen that

$$P = \alpha. \qquad (3)$$

P is the probability of an event, say E, which we shall describe in some detail. Let us denote generally the probability of any event a by $P\{a\}$. $P\{a \mid b\}$ will denote the probability of an event, a, calculated under the assumption that another event, b, has already occurred.

Now

$P = P\{E\} =$ the probability that {either $(x = x_1)$ and then $\theta_1(1) \leqq \theta \leqq \theta_2(1)$

$$\text{or } (x = x_2) \text{ and then } \theta_1(2) \leqq \theta \leqq \theta_2(2)$$

$$\text{or } (x = x_n) \quad \text{"} \quad \text{"} \quad \theta_1(n) \leqq \theta \leqq \theta_2(n)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\}$$

$$= P\{x = x_1\}P\{\theta_1(1) \leqq \theta \leqq \theta_2(1) \mid (x = x_1)\}$$

$$+ P\{x = x_2\}P\{\theta_1(2) \leqq \theta \leqq \theta_2(2) \mid (x = x_2)\}$$

$$+ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$= \sum_{n=1}^{\infty} P\{x = x_n\}P\{\theta_1(n) \leqq \theta \leqq \theta_2(n) \mid (x = x_n)\} = \alpha \ldots\ldots\ldots\ldots\ldots \quad (4)$$

The calculation of the probability P in the above form is not convenient, as both multipliers in each term of the sum in (4) depend upon the unknown probability function *a priori* of θ. Therefore we shall present P in another form, giving to the event E a geometrical interpretation. Let us denote by CB the set of all confidence intervals $(\theta_1(n), \theta_2(n))$, as marked on the plane of x and θ. Thus CB will be composed of points with co-ordinates x and θ, where

$$\begin{matrix} x = x_n \\ \theta_1(n) \leqq \theta \leqq \theta_2(n) \end{matrix} \quad n = 1, 2, \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (5)$$

The set CB will be called the confidence belt.

Denote by A any point of the plane of x and θ, having any values for its co-ordinates.

It is easily seen that the event, which we denote by E, and the probability of which is $P = \alpha$, consists in the point A belonging to the confidence belt CB. In fact the event E occurs if and only if the co-ordinates of A fulfil the conditions (5). But just these conditions define the points belonging to CB.

The above circumstance allows us to calculate P by means of a formula which discloses its connection with $p(x \mid \theta)$.

Fix any possible value of $\theta = \theta'$ and draw the straight line LL the points of which have just this fixed value θ' for their ordinates. The line LL will cut some of the confidence intervals. Denote by $X(\theta')$ the set of points of intersection, and by $\phi(\theta)$ the unknown frequency function of θ. The set $X(\theta)$ will be called the set of acceptance corresponding to the specified value of θ.

The function $\phi(\theta)$ may be continuous or not. So may be $p(x \mid \theta)$ considered as a function of θ. These cases may be treated together if we agree that $\sum_{\theta} F(\theta)$ will denote either the sum or the integral of $F(\theta)$ extending over all values of θ, whenever $F(\theta)$ is integrable.

Using this notation we may write

$$P = P\{E\} = \sum_{\theta} \left\{ \phi(\theta) \sum_{X(\theta)} (p(x \mid \theta)) \right\} \quad \ldots\ldots\ldots\ldots\ldots (6)$$

where $\sum_{X(\theta)}$ denotes the summation over all values of x belonging to $X(\theta)$.

From the formula (6) may be deduced the following important proposition. *The probability P may possess a constant value α, independent of the properties of the unknown function $\phi(\theta)$, if and only if for each θ*

$$\sum_{X(\theta)} (p(x \mid \theta)) = \alpha. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (7)$$

The condition (7) is obviously sufficient to have $P = \alpha$. In fact, if it is satisfied, then we should get from (6)

$$P = \alpha \sum_{\theta} (\phi(\theta)) = \alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (8)$$

since $\sum_{\theta} (\phi(\theta)) = 1$ whatever the frequency distribution of θ. It is equally easy to see that the condition (7) is necessary for having $P = \alpha$ whatever the function $\phi(\theta)$. For suppose that for $\theta = \theta_1$ we have

$$\sum_{X(\theta_1)} (p(x \mid \theta_1)) = \beta \neq \alpha \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (9)$$

Then if it happens, that

$$\phi(\theta_1) = 1 \qquad \text{for } \theta = \theta_1 \tag{10}$$

and

$$\phi(\theta) = 0 \qquad \text{for } \theta \neq \theta_1 \tag{11}$$

the only term in the sum $\sum_{\theta}$ which is different from zero will be that corresponding to $\theta = \theta_1$ and the formula (6) will reduce to

$$P = \sum_{X(\theta_1)} (p(x \mid \theta_1)) = \beta \neq \alpha. \tag{12}$$

The original question, whether the solution of the form (a) is possible when the variate x is discontinuous is thus put in the following form: is it possible to define for every possible value of θ a set of acceptance $X(\theta)$ such that the equation (7) holds good?

The answer is: in some cases it may be possible, but this depends upon the nature of the function $p(x \mid \theta)$. It is very easy to *invent* functions $p(x \mid \theta)$ for which the equation (7) for a definite value of α holds good, and we may even fix in advance the sets of acceptance $X(\theta)$. However the important question is not whether there may exist elaborately invented cases of discontinuous distributions where the solution (a) exists, but rather whether this solution exists always, or at least whether it exists frequently and in cases which are practically important.

This question must be answered in the negative on the basis of the following example concerning the most important of the discontinuous distributions, the Binomial.

In fact it will be seen below that if x is a variate following the binomial frequency law, then whatever the arrangement of the sets of acceptance $X(\theta)$, corresponding to different values of θ, the left hand side of the equation (7) cannot be constantly equal to the confidence coefficient $\alpha < 1$. It will follow that in the case of the binomial distribution, the solution of the problem (a) is impossible.

To prove this we shall consider the variate, x, following the binomial frequency law. That is to say we shall assume that x may have values $0, 1, 2, \cdots n$, and that

$$p(x \mid \theta) = \frac{n!}{x!(n-x)!}\,\theta^x(1-\theta)^{(n-x)} \tag{13}$$

while $0 < \theta < 1$. Since the set of possible values which x may have is finite, therefore the set of all confidence intervals must be finite also. It follows that there is possible only a finite number of sets of acceptance $X(\theta)$. Therefore there must be at least one set of acceptance, say X^0, which will be common to an infinite number of values of θ, say $\theta_1, \theta_2, \cdots \theta_n, \cdots$ so that for each it will be $X(\theta_n) = X^0$.

Now

$$\sum_{X(\theta_n)} (p(x \mid \theta_n)) \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

for all these values of $\theta = \theta_n$ will be *the same* polynomial in θ of the order n. If it has the same value α for a number of values of θ exceeding n, it means that this polynomial is an absolute constant. Therefore if it were possible to give a solution of the type (a) in the case of the binomial distribution, it would be possible to construct a sum (14), the terms of which are all different and have the form (13), and such that after all possible reductions and simplifications all terms involving θ would cancel and we should be left only with one constant term $\alpha<1$. This, however, is impossible, since the only term of the form (13) which involves a constant, is the term corresponding to $x = 0$

$$p(0 \mid \theta) = (1-\theta)^n = 1 - n\theta + \frac{n(n-1)}{2}\theta^2 \ \ldots\ldots\ldots\ldots (15)$$

and then this constant is 1. Other terms of the form (13) involve θ^x as a multiplier. Therefore there exists only one sum of the form (14) which is an absolute constant, but this includes all the terms (13)

$$\sum_{x=0}^{n} (p(x \mid \theta)) = 1 \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16)$$

and thus is of no value. It follows that whatever the sets of acceptance $X(\theta)$

the corresponding sum (14) will have values varying with the value of θ and hence the solution of the type (a) in the case of the binomial does not exist.

This, I think, gives the solution of the question raised by Professor Fisher. It is clear also that whenever the solution of the type (a) exists, it may be found by a suitable choice of sets of acceptance, and thus by the method explained in my earlier paper.

I should like now to raise another question. Past experience shows that the general problem of estimation may be formulated in different ways. The form of this problem as it appears in Bayes theorem, required for its solution the knowledge of the probabilities *a priori*.

The form of the same problem treated by R. A. Fisher in his theory of estimation was solved in terms of a new conception, that of likelihood.

The problem of estimation in its form of confidence intervals stands entirely within the bounds of the theory of probability, without involving any conception not already inherent in this theory. In the case of continuous distribution the problem also allows the solution (a) entirely independent of the probabilities *a priori*. Now it is shown that the necessity of the solution (b) is bound up with the nature of the problem if the distributions are discontinuous.

My question is: is it possible to formulate the problem of estimation in a fourth form, leading to a solution which (1) stands entirely on the grounds of the classical theory of probability, and (2) is not depending upon the probabilities *a priori*—whatever the conditions of the problem?

ANALYSIS OF VARIANCE CONSIDERED AS AN APPLICATION OF SIMPLE ERROR THEORY

By Walter A. Hendricks

The need for an elementary presentation of the methods of analysis of variance has been recognized by many investigators in various fields of research. A recent monograph by Snedecor (1934) is undoubtedly the most comprehensive attempt to satisfy this need which has appeared in the literature relating to the subject. Snedecor's treatment of the subject consists largely of the presentation of a number of standard types of problems to which the methods of analysis of variance are applicable, directions for performing the necessary computations, and a discussion of the conclusions which may be drawn from the data on the basis of the analysis.

In the opinion of the author of this paper, an elementary presentation of some of the theoretical considerations upon which the methods of analysis of variance are based would also be of some value. The methods of analysis of variance, as given by Fisher (1932), are presented as a natural consequence of intraclass correlation theory. However, the essential concepts may be presented in a more comprehensible form by the use of simple error theory.

It seems appropriate to begin such a presentation with a definition of variance. If we have an infinite number of measurements of the same quantity, the variance of a single measurement is defined as the arithmetic mean of the squares of the errors of those measurements. In actual practice, an infinite number of measurements can never be obtained. We have instead a sample of n measurements, $x_1, x_2, \cdots x_n$, from which the variance of a single measurement may be estimated. By referring to any text on the method of least squares, it may be verified that the best estimate, S^2, of the variance of a single measurement which can be obtained from a sample of n measurements is given by the equation:

$$S^2 = \frac{1}{n-1} \sum_{i}^{n} (x_i - m)^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

in which m represents the arithmetic mean of the n measurements. The quantity, $n - 1$, in the terminology of analysis of variance, is designated as the number of degrees of freedom available for estimating S^2.

It is often necessary to estimate S^2 from a number of different samples of measurements. In such cases, the best estimate of S^2 is obtained by calculating the weighted mean of the variances estimated from the individual samples, each variance being weighted by the number of degrees of freedom which were avail-

able for its estimation. The number of degrees of freedom upon which such an estimate of S^2 is based is given by the sum of these weights. Such an estimate of the variance of a single measurement is often designated as the variance "within samples."

In one of the simpler applications of analysis of variance, a number of samples of measurements are available, and the investigator is required to determine whether the magnitude of the quantity measured varied from sample to sample or whether all of the measurements may be regarded as having been made upon a quantity of the same magnitude.

An estimate, S^2, of the variance within samples may be obtained. Since S^2 is an estimate of the variance of a single measurement, the variance, S_i^2, of the arithmetic mean, m_i, of the measurements in any one sample is given by the equation:

$$S_i^2 = \frac{S^2}{n_i} \quad \text{(2)}$$

in which n_i represents the number of measurements in the sample. Let there be r samples. Then another estimate, $S_i'^2$, of the variance of the mean, m_i, may be obtained from the observed distribution of the means, $m_1, m_2, \cdots m_r$, by the use of the formula for calculating the variance of a weighted observation as given in texts on the method of least squares:

$$S_i'^2 = \frac{1}{n_i(r-1)}[n_1(m_1 - m)^2 + n_2(m_2 - m)^2 + \cdots + n_r(m_r - m)^2] \quad \text{(3)}$$

in which:

$$m = \frac{n_1 m_1 + n_2 m_2 + \cdots + n_r m_r}{n_1 + n_2 + \cdots + n_r} \quad \text{(4)}$$

Equations (2) and (3) yield two estimates of the variance of the mean, m_i. It is apparent that these two estimates will be equal, within the limits of sampling fluctuations, if all of the measurements in the r samples were made upon a quantity of the same magnitude. If the magnitude of the quantity measured varied from sample to sample, $S_i'^2$ will be greater than S_i^2. However, in actual practice, the two estimates of the variance of a particular mean are not compared directly. An equivalent comparison is made between two estimates of the variance of a single measurement. The first of these is nothing more than the variance within samples discussed earlier in this paper. The second estimate, which may be designated by S'^2, is the value which would have to be substituted for S^2 in equation (2) in order to make S_i^2 equal to the value given for $S_i'^2$ by equation (3). It is quite apparent that S'^2 may be found by the use of the equation:

$$S'^2 = \frac{1}{r-1}[n_1(m_1 - m)^2 + n_2(m_2 - m)^2 + \cdots + n_r(m_r - m)^2]. \quad \text{(5)}$$

S'^2 is often designated as the variance "between samples." A comparison of S'^2 with S^2 is obviously equivalent to a comparison of $S_i'^2$ with S_i^2.

If S'^2 is greater than S^2, a statistic, z, may be calculated:

$$z = \frac{1}{2} \log_e \frac{S'^2}{S^2} \qquad (6)$$

This statistic serves as a useful comparison between S'^2 and S^2 since its sampling distribution is known if all of the measurements comprising the data under investigation were made upon a quantity of the same magnitude. The distribution of z, under these conditions, is given by an equation of the form:

$$df = \frac{ke^{n_1 z}}{(n_1 e^{2z} + n_2)^{\frac{1}{2}(n_1+n_2)}} dz \qquad (7)$$

in which n_1 represents the number of degrees of freedom available for estimating S'^2, and n_2 represents the number of degrees of freedom available for estimating S^2. It is apparent from equation (5) that $r - 1$ degrees of freedom are available for the estimation of S'^2 in the particular problem under discussion.

When any estimate of the variance of a single measurement is multiplied by the number of degrees of freedom available for making that estimate, the resulting product is known as a "sum of squares." The additive property of the sums of squares and the degrees of freedom contributes much to the elegance of the scheme of analysis just presented and is of considerable practical importance in problems of a type to be discussed later in this paper. In the case of the problem discussed above, the additive property of the sums of squares provides that the sum of the "sum of squares between samples" and the "sum of squares within samples" is equal to the sum of the squares of the deviations of all of the measurements from their arithmetic mean. The additive property of the degrees of freedom provides that the sum of the "degrees of freedom between samples" and the "degrees of freedom within samples" is equal to the "total degrees of freedom" which is nothing more than the total number of measurements diminished by unity.

The methods of analysis presented above may be applied to any study of the effects of a number of experimental treatments of the same kind upon the magnitude of a measurable quantity. If experimental treatments of more than one kind are imposed simultaneously, the effects of each may be studied by modifications of those methods. The discussion of those modifications, about to be presented in this paper, is limited to data which may be classified in an "$r \times s$" table, i.e., to studies of the effects of only two kinds of experimental treatments. More complex problems may be treated by simple extensions of the methods presented.

Consider an "$r \times s$" table composed of rs cells, each of which contains a number of measurements of some quantity. The magnitude of the quantity measured may vary from cell to cell, but the essential conditions under which the measurements were made must be the same for all cells. It is also under-

stood that no cell may be empty. Table 1 is an example of such a table. The individual measurements have not been represented. Only the number of measurements, n_{ij}, in each cell and the arithmetic mean, m_{ij}, of those measurements have been indicated. The arguments, a_i, represent r experimental treatments of one kind, and the arguments, b_j, represent s experimental treatments of another kind. The problem to be solved is to ascertain whether or not the differences among the experimental treatments of each kind had any effect on the magnitude of the quantity measured.

TABLE 1

Example of an "$r \times s$" Table Showing Only the Number of Measurements in Each Cell and the Arithmetic Mean of Those Measurements

	b_1	b_2	b_3	b_4		b_s
a_1	m_{11} n_{11}	m_{12} n_{12}	m_{13} n_{13}	m_{14} n_{14}		m_{1s} n_{1s}
a_2	m_{21} n_{21}	m_{22} n_{22}	m_{23} n_{23}	m_{24} n_{24}		m_{2s} n_{2s}
a_3	m_{31} n_{31}	m_{32} n_{32}	m_{33} n_{33}	m_{34} n_{34}		m_{3s} n_{3s}
a_r	m_{r1} n_{r1}	m_{r2} n_{r2}	m_{r3} n_{r3}	m_{r4} n_{r4}		m_{rs} n_{rs}

If each cell contains the same number of measurements, the effects of the experimental treatments indicated by the arguments, a_i, may be studied by comparing the variance "between rows" with the variance "within cells." The variance between rows may be calculated by regarding the r rows as r samples of measurements and applying an equation of the same form as equation (5). The variance within cells may be obtained by calculating the variance of a single measurement from the data in each cell separately and taking the mean of the resulting values. The effects of the experimental treatments indicated by the arguments, b_j, may be studied by comparing the variance "between columns" with the variance "within cells."

If the degrees of freedom between rows, between columns, and within cells are added, the sum will be less than the total number of degrees of freedom in the table. If the corresponding sums of squares are added, the sum is likely to be less than the total sum of squares. The differences are due to what is customarily designated as "interaction between rows and columns." The

more descriptive term, "differential response," is sometimes used to designate the same factor. The nature of this factor may be investigated by considering the effects of the experimental treatments, b_j, in each row of Table 1.

The data in each cell of Table 1 may be regarded as a sample of measurements. Therefore, the data in any row may be regarded as a set of s samples of measurements. By applying an equation of the same form as equation (5) to the data in any row, an estimate of the variance of a single measurement is obtained from the observed distribution of the means of the cells in that row. By calculating the arithmetic mean of the estimates for the r rows, an estimate of the variance of a single measurement is obtained from $r(s - 1)$ degrees of freedom. This estimate may be designated as the variance "between cells in the same row."

The variance between cells in the same row measures the average effect of differences among the experimental treatments, b_j, in individual rows. The variance between columns, which was discussed earlier in this paper, is calculated from $s - 1$ degrees of freedom and measures the effect of differences among the treatments, b_j, on the assumption that the effect of any one treatment upon the magnitude of the quantity measured was constant for every row. The number of degrees of freedom assignable to differential response of the various rows to the treatments, b_j, is $r(s - 1) - (s - 1)$ or $(r - 1)(s - 1)$. The sum of squares due to differential response is given by the difference between the sum of squares between cells in the same row and the sum of squares between columns. These relations follow from the additive property of degrees of freedom and sums of squares.

It may be observed that precisely the same results would be obtained by considering the effects of the treatments, a_i, in the various columns of Table 1. The degrees of freedom and sum of squares due to differential response of the various columns to the treatments, a_i, would be exactly equal to the corresponding values obtained for the differential response of the various rows to the treatments, b_j.

Up to this point the discussion has been concerned only with the special case in which each cell of Table 1 contains the same number of measurements. As a matter of fact, the methods given for the analysis of such data will yield correct results when applied to any "$r \times s$" table in which the numbers of measurements in the cells in every row are proportional to the corresponding marginal totals for the columns, and the numbers of measurements in the cells in every column are proportional to the corresponding marginal totals for the rows.

When the numbers of measurements in the various cells do not satisfy the above condition of proportionality, the distributions of the means of the rows and columns may be distorted, and, consequently, the methods of analysis described above may yield incorrect results. Efficient methods of analyzing such data have been presented by Yates (1933). A comprehensive discussion of these methods is considerably beyond the scope of this paper. One method,

described very briefly by Yates (1933) and designated as the "method of weighted squares of means," appealed to the author as being particularly valuable for practical work. No detailed discussion of the method seems to be available in the literature. Therefore, the following presentation may be of some interest.

Consider the experimental treatments represented by the arguments, a_i, in Table 1. It is necessary to find an average value for the magnitude of the quantity measured for each row of Table 1. However, this average must be of such a type that its value will not be distorted by the unequal numbers of measurements in the various cells. The unweighted arithmetic mean of the means of the cells in the row seems to be the logical average to use since, within the limits of sampling fluctuations, the value of this average will be identical with the value which would have been obtained if each cell had contained the same number of measurements. The averages for the r rows are:

$$m_{a} = \frac{1}{s}(m_{11} + m_{12} + \cdots + m_{1s})$$

$$m_{a_2} = \frac{1}{s}(m_{21} + m_{22} + \cdots + m_{2s})$$

$$m_{a_r} = \frac{1}{s}(m_{r1} + m_{r2} + \cdots + m_{rs}). \qquad (8)$$

By the law of propagation of error, the variance of any one of these unweighted means is given by the equation:

$$S^2_{a_i} = \frac{1}{s^2}(S^2_{i1} + S^2_{i2} + \cdots + S^2_{is}) \qquad (9)$$

in which $S^2_{a_i}$ is the variance of m_{a_i}, and $S^2_{i1}, S^2_{i2}, \cdots, S^2_{is}$ are the variances of $m_{i1}, m_{i2}, \cdots, m_{is}$, respectively. If S^2 represents the variance of a single measurement, equation (9) may be written in the form:

$$S^2_{a_i} = \left(\frac{1}{n_{i1}} + \frac{1}{n_{i2}} + \cdots + \frac{1}{n_{is}}\right)\frac{S^2}{s^2} \qquad (10)$$

The value of S^2 may be estimated from the individual measurements in the various cells. S^2 is nothing more than the variance within cells, as customarily calculated, and may be estimated from the $N - rs$ degrees of freedom within cells, in which N represents the total number of measurements in Table 1.

The variance of a single measurement may also be estimated from the observed distribution of the means of the type, m_{a_i}. These means are not of equal weight. Therefore, in order to find the variance of any one of them, it is first necessary to calculate the weighted mean of the r individual means. Since the weight of an arithmetic mean is inversely proportional to its variance, it is evident from

an inspection of equation (10) that the weight, p_{a_i}, of a mean, m_{a_i}, may be found from the equation:

$$\frac{1}{p_{a_i}} = \frac{1}{n_{i1}} + \frac{1}{n_{i2}} + \cdots + \frac{1}{n_{is}} \qquad \text{(11)}$$

The weighted mean, m_a, may then be found:

$$m_a = \frac{p_{a_1}m_{a_1} + p_{a_2}m_{a_2} + \cdots + p_{a_r}m_{a_r}}{p_{a_1} + p_{a_2} + \cdots + p_{a_r}} \qquad \text{(12)}$$

The variance $S'^2_{a_i}$, of any mean, m_{a_i}, as estimated from the observed distribution of means of this type, is given by:

$$S'^2_{a_i} = \frac{1}{p_{a_i}(r-1)}[p_{a_1}(m_{a_1} - m_a)^2 + p_{a_2}(m_{a_2} - m_a)^2 + \cdots + p_{a_r}(m_{a_r} - m_a)^2]. \qquad \text{(13)}$$

By substituting $S'^2_{a_i}$ for $S^2_{a_i}$, and S^2_a for S^2, in equation (10) and solving the resulting equation for S^2_a, an estimate, S^2_a, of the variance of a single measurement is obtained from the observed distribution of means of the type, m_{a_i}. It is evident that, after making the indicated substitutions, equation (10) reduces to the form:

$$S^2_a = \frac{s^2}{r-1}[p_{a_1}(m_{a_1} - m_a)^2 + p_{a_2}(m_{a_2} - m_a)^2 + \cdots + p_{a_r}(m_{a_r} - m_a)^2] \qquad \text{(14)}$$

It is interesting to observe that, if the numbers of measurements in the respective cells were equal, equation (14) would reduce to the formula for calculating the variance "between rows" as customarily applied in analysis of variance.

The two estimates, S^2 and S^2_a, of the variance of a single measurement may be compared in the usual manner by taking one-half of the natural logarithm of the ratio of the larger estimate to the smaller and making use of the tables of the values of "z" given by Fisher (1932). When using these tables, it is important to remember that S^2_a was estimated from $r - 1$ degrees of freedom.

The method of analysis just described may be employed to study the effects of differences among the experimental treatments indicated by the arguments, b_j, on the magnitude of the quantity measured. The unweighted means for the s columns are:

$$m_{b_1} = \frac{1}{r}(m_{11} + m_{21} + \cdots + m_{r1})$$

$$m_{b_2} = \frac{1}{r}(m_{12} + m_{22} + \cdots + m_{r2})$$

$$m_{b_s} = \frac{1}{r}(m_{1s} + m_{2s} + \cdots + m_{rs}) \qquad \text{(15)}$$

The weight, p_{b_j}, of a mean of the type, m_{b_j}, may be found from the relation:

$$\frac{1}{p_{b_j}} = \frac{1}{n_{1j}} + \frac{1}{n_{2j}} + \cdots + \frac{1}{n_{rj}} \qquad (16)$$

A weighted mean, m_b, may be calculated:

$$m_b = \frac{p_{b_1}m_{b_1} + p_{b_2}m_{b_2} + \cdots + p_{b_s}m_{b_s}}{p_{b_1} + p_{b_2} + \cdots + p_{b_s}} \qquad (17)$$

An estimate, S_b^2, of the variance of a single measurement may be obtained from the observed distribution of means of the type, m_{b_j}, by the use of the equation:

$$S_b^2 = \frac{r^2}{s-1}\left[p_{b_1}(m_{b_1} - m_b)^2 + p_{b_2}(m_{b_2} - m_b)^2 + \cdots + p_{b_s}(m_{b_s} - m_b)^2\right] \qquad (18)$$

S_b^2 may be compared with S^2 in the usual manner.

If it is necessary to study the "interaction between rows and columns," the effects of the experimental treatments, b_j, may be studied for each individual row of Table 1. Consider the distribution of the means of the cells in a row designated by the argument, a_i. The weight of any one of these means is equal to the number of measurements in the cell. A weighted mean, m'_{a_i}, of the s means of cells in the row may be calculated:

$$m'_{a_i} = \frac{n_{i1}m_{i1} + n_{i2}m_{i2} + \cdots + n_{is}m_{is}}{n_{i1} + n_{i2} + \cdots + n_{is}} \qquad (19)$$

The variance, S'^2_{ij}, of the mean, m_{ij}, for any cell in the given row, as estimated from the observed distribution of means of this type, may be obtained from the equation:

$$S'^2_{ij} = \frac{1}{n_{ij}(s-1)}\left[n_{i1}(m_{i1} - m'_{a_i})^2 + n_{i2}(m_{i2} - m'_{a_i})^2 + \cdots + n_{is}(m_{is} - m'_{a_i})^2\right] \qquad (20)$$

The variance, S^2_{ij}, of the same mean, as estimated from the distribution of the individual measurements in the cell, may be obtained from the equation:

$$S^2_{ij} = \frac{S^2}{n_{ij}} \qquad (21)$$

By substituting S'^2_{ij} for S^2_{ij}, and $S^2_{a_ib}$ for S^2, in equation (21) and solving the resulting equation for $S^2_{a_ib}$, an estimate, $S^2_{a_ib}$, of the variance of a single measurement is obtained from the observed distribution of the means of the cells in the given row. After making the indicated substitutions, equation (21) reduces to the form:

$$S^2_{a_ib} = \frac{1}{s-1}\left[n_{i1}(m_{i1} - m'_{a_i})^2 + n_{i2}(m_{i2} - m'_{a_i})^2 + \cdots + n_{is}(m_{is} - m'_{a_i})^2\right] \qquad (22)$$

Such an estimate, $S^2_{a_i b}$, of the variance of a single measurement may be obtained for each of the r rows in Table 1. By calculating the average, S^2_{ab}, of the variances of the type, $S^2_{a_i b}$, an estimate, S^2_{ab}, of the variance of a single measurement may be obtained from the $r(s - 1)$ degrees of freedom between cells in the same row:

$$S^2_{ab} = \frac{1}{r(s-1)} \sum_{i=1}^{r} [n_{i1}(m_{i1} - m'_{a_i})^2 + n_{i2}(m_{i2} - m'_{a_i})^2 + \cdots + n_{is}(m_{is} - m'_{a_i})^2] \;\ldots\ldots (23)$$

Equation (23) is identical with the formula for calculating the variance between cells in the same row as ordinarily applied in analysis of variance. This result is a direct consequence of the fact that the unequal numbers of measurements in the various cells had no distorting effect on the arithmetic means for individual cells.

The presence or absence of interaction may be verified by comparing S^2_{ab} with S^2_b. In general, the actual variance due to interaction can not be obtained by the "weighted squares of means" method, for the various sums of squares do not possess the additive property when the analysis is made in this way. However, the comparison suggested above will yield sufficient information for most practical purposes.

For the special case in which r or s is equal to 2, the actual variance due to interaction may be calculated. Suppose $r = 2$ in Table 1. The following method, suggested by Yates (1933), yields an estimate of the variance due to interaction from a consideration of the differences, d_j, between the means of the two cells in each column:

$$\begin{aligned} d_1 &= m_{11} - m_{21} \\ d_2 &= m_{12} - m_{22} \\ &\vdots \qquad \vdots \\ d_s &= m_{1s} - m_{2s} \;\ldots\ldots (24) \end{aligned}$$

The variance, $S^2_{d_j}$, of any difference, d_j, is given by the equation:

$$S^2_{d_j} = \left(\frac{1}{n_{1_j}} + \frac{1}{n_{2_j}}\right) S^2 \;\ldots\ldots (25)$$

The weight, p_j, of the difference, d_j, is given by the equation:

$$\frac{1}{p_j} = \frac{1}{n_{1j}} + \frac{1}{n_{2j}} \;\ldots\ldots (26)$$

The variance of the difference, d_j, as estimated from the observed distribution of differences, is given by the equation:

$$S'^2_{d_j} = \frac{1}{p_j(s-1)} [p_1(d_1 - d)^2 + p_2(d_2 - d)^2 + \cdots + p_s(d_s - d)^2] \;\ldots (27)$$

in which:

$$d = \frac{p_1 d_1 + p_2 d_2 + \cdots + p_s d_s}{p_1 + p_2 + \cdots + p_s} \quad \ldots\ldots\ldots\ldots\ldots\ldots (28)$$

By means of these relations, an estimate, S_d^2, of the variance of a single measurement may be obtained from the observed distribution of the differences of the type, d_j. This estimate represents the variance due to interaction and may be obtained from the equation:

$$S_d^2 = \frac{1}{s-1}\left[p_1(d_1 - d)^2 + p_2(d_2 - d)^2 + \cdots + p_s(d_s - d)^2\right] \quad \ldots\ldots (29)$$

It is quite apparent that $s - 1$ degrees of freedom are available for the estimation of the variance due to interaction in this particular example.

REFERENCES

FISHER, R. A., 1932. Statistical Methods for Research Workers, 4th edition. Edinburgh and London: Oliver and Boyd

SNEDECOR, GEORGE W., 1934. Calculation and Interpretation of Analysis of Variance and Covariance. Ames, Iowa: Collegiate Press.

YATES, F., 1933. The principles of orthogonality and confounding in replicated experiments. Jour. Agr. Sci., 23: 108–145.

BUREAU OF ANIMAL INDUSTRY,
U. S. DEPARTMENT OF AGRICULTURE,
WASHINGTON, D. C.

NOTE ON THE DISTRIBUTIONS OF THE STANDARD DEVIATIONS AND SECOND MOMENTS OF SAMPLES FROM A GRAM-CHARLIER POPULATION

By G. A. Baker

T. N. Thiele in his "Theory of Observations" makes the following statement with regard to the distributions of the higher half-invariants in samples of n: "Not even for μ_2 have I discovered the general law of errors."[1] The purpose of this paper is to shed some light on the distribution of μ_2 and to give the distribution of second moments about a fixed point when the sampled population can be represented by a Gram-Charlier series.

The distribution of the second moments about a fixed point of samples is given in complete generality. It is known that if the sampled population is normal there is a simple relation between the distribution of the standard deviations of samples of n and the distribution of the second moments of the samples about the mean of the population. It was thought that such a relation might exist in case the sampled population could be represented by a Gram-Charlier series. Such is not the case. Again, it was thought that by obtaining the distribution of the standard deviations for samples of 2, 3, 4, $\cdots$ it might be possible to deduce empirically a general law of distribution. This proved an unfruitful line of investigation but required so much labor that the results should be reported to save others time and energy.

First, suppose that a population may be represented as

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_5\varphi_5(x) + \cdots \tag{1}$$

where

$$\varphi_i(x) = \frac{d^i(e^{-\frac{1}{2}x^2})}{dx^i} .$$

Then applying Theorem II of the author's paper on "Random Sampling from Non-Homogeneous Populations"[2] we deduce at once the following theorem.

Theorem I. The distribution of the second moments about the origin of (1) of samples of n drawn at random from a population represented by (1) is precisely the same as the distribution of the second moments about the same point of samples of n drawn from a population represented by the first term of (1), that is a normal population, and is proportional to $x^{-\frac{\quad}{2}} e^{-\frac{1}{2}x}$ (loc. cit.)

[1] Thiele, T. N., "The Theory of Observations," reprinted in the *Annals of Mathematical Statistics*, Vol. 2, No., 2, May, 1931, p. 208.

[2] *Metron*, Vol. 8, No. 3, Feb. 28, 1930.

This is not so surprising as it may seem at first if it is remembered that the odd subscript terms of a Gram-Charlier series slice off frequencies on one side of the mean of $a_0\varphi_0(x)$ and add them onto the other side in the same manner.

If we suppose that a population is given as

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_4\varphi_4(x) + \cdots \tag{2}$$

in the same manner we get the following theorem.

THEOREM II. The distribution of the second moments measured from the origin of (2) of samples of n drawn at random from (2) will be a combination of distributions of the type of Theorem I with only even subscript terms contributing anything. The variations in the component distributions will consist of differences in the constant factors and the exponent of x, the estimate of the second moment. The lowest exponent will be $\frac{n-2}{2}$.

For instance, if

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_4\varphi_4(x) \tag{3}$$

and $n = 2$, the estimates of the second moment will be distributed as proportional to

$$e^{-\frac{1}{2}x}\left[(a_0+3)^2 - 12a_4(a_0+3)x + (36a_4^2 + 6a_0a_4 + 18a_4)\frac{x^2}{2!}\right.$$
$$\left. 36a_4^2\frac{x^3}{3!} + 9a_4^2\frac{x^4}{4!}\right].$$

Thus, it can be said that we know the distribution of the second moments of samples about a fixed point if the sampled population is of the Gram-Charlier type in the sense that given the number of terms necessary for an adequate representation and the number in the samples we can write down the desired distribution. However, this is not a simple matter. Further, if some relation existed between the distributions of the second moments about a fixed point and the standard deviations of the samples we would know the latter distribution also. Such a relation is not apparent for samples of 2 and 3.

Let us investigate the correlation surfaces of the means and standard deviations of samples of 2 and 3 drawn at random from a population represented by the first few terms of a Gram-Charlier series after the method of Dr. A. T. Craig.[3] The distributions of the standard deviations can then be obtained immediately by integration.

Suppose that

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) + a_4\varphi_4(x) \tag{4}$$

[3] *Annals of Mathematical Statistics*, Vol. 3, No. 2, May, 1932, pp. 126–140.

and that we are considering samples of 2. The probability of the concurrence of x_1 and x_2 is

$$f(x_1)f(x_2) \tag{5}$$

and

$$\begin{aligned} x_1 &= -s + x \\ x_2 &= s + x \end{aligned} \tag{6}$$

where s is the standard deviation and x is the mean of a sample of 2. By means of (6), (5) becomes

$$\begin{aligned} &e^{-(s^2+x^2)}[a_0^2 + a_0a_3(-6s^2x - 2x^3 + 6x) \\ &+ a_0a_4(2s^4 + 12s^2x^2 - 12s^2 - 12x^2 + 6) \\ &+ a_3^2(-s^6 + 3s^4x^2 + 6s^4 - 3s^2x^4 - 9s^2 + 9x^2 - 6x^4 + x^6) \\ &+ a_3a_4(2s^6 - 6s^4x^3 - 6s^4x + 6s^2x^5 - 12s^2x^3 + 18s^2x - 2x^7 \\ &\qquad + 18x^5 - 42x^3 + 18x) \\ &+ a_4^2(s^8 - 4s^6x^2 - 12s^6 + 6s^4x^4 + 12s^4x^2 + 42s^4 - 4s^2x^6 \\ &\qquad + 12s^2x^4 - 36s^2x^2 - 36s^2 + x^8 - 12x^6 + 42x^4 - 36x^2 + 9)]. \end{aligned} \tag{7}$$

To find the distribution of s we must integrate from $-\infty$ to ∞ with respect to x. Thus, (8) is obtained.

$$\begin{aligned} &\sqrt{\pi}\, e^{-s^2}\left[a_0^2 + a_0a_4(2s^4 - 6s^2) + a_3^2\left(-s^6 + \frac{15}{2}s^4 - \frac{45}{4}s^2 + \frac{15}{8}\right)\right. \\ &\qquad \left. + 2a_3a_4s^6 + a_4^2\left(s^8 - 14s^6 + \frac{105}{2}s^4 - \frac{105}{2}s^2 + \frac{105}{2}\right)\right]. \end{aligned} \tag{8}$$

If we retain only two terms of (3), i.e. use

$$f(x) = a_0\varphi_0(x) + a_3\varphi_3(x) \tag{9}$$

and consider samples of 3 we obtain as the correlation surface of x and s

$$\begin{aligned} &\frac{18\pi}{\sqrt{3}} s e^{-\frac{1}{2}(3x^2+3s^2)}\left[a_0^3 - \frac{a_0^2a_3}{4}(-40x^3 + 24xs^2 - 24x)\right. \\ &\qquad + \frac{a_0a_3^2}{64}(-84s^6 + 525x^2s^4 - 2752x^4s^2 \\ &\qquad + 576s^4 - 1008x^2s^2 - 288s^2 - 5586x^6 + 270x^4 - 1728x^2) \\ &\qquad + \frac{a_3^3}{64}(28s^6 - 6189x^2s^4 - 28x^4s^2 - 629x^6 + 288s^4 + 1344x^4 \\ &\qquad \left. + 4608x^2s^2 - 288s^2 + 729x^2)\right]. \end{aligned} \tag{10}$$

The distribution of s can be obtained as before. The processes involved in obtaining (7) and (10) are so complicated that the general rule for writing the distribution of s is not apparent. Also, the relation of the distributions of s to the corresponding distributions of the second moments about a fixed point is not apparent.

In summary, the general distributions of the second moments about a fixed point of samples from a population represented by a definite number of terms of a Gram-Charlier series and the distributions of the standard deviations of samples of 2 and 3 from the same type of population are given and compared. No apparent relation exists between them.

ON THE FINITE DIFFERENCES OF A POLYNOMIAL

By I. H. Barkey

In this paper an apparently new and convenient method of finding the successive finite differences of a polynomial is considered. If operationally

$$\phi(u + r_1 r_2) = E^{r_1 r_2}\,\phi(u) = (1 + \Delta r_1)^{r_2}\,\phi(u)$$

then for any polynomial $f(x)$ of degree "n"

$$\begin{aligned}
f(x) &= p_0 x^n + p_1 x^{n-1} + \cdots + p_n \\
&= p_0(x + a)^n + q_{11}(x + a)^{n-1} + \cdots + q_{1n} \\
E^a f(x) &= p_0(x + a)^n + p_1(x + a)^{n-1} + \cdots + p_n \\
\Delta_a f(x) &= (p_1 - q_{11})(x + a)^{n-1} + (p_2 - q_{12})(x + a)^{n-2} + \cdots + (p_n - q_{1n}).
\end{aligned}$$

Similarly, if $f_1(x) = \Delta a f(x)$, then

$$\begin{aligned}
f_1(x) &= (p_1 - q_{11})(x + 2a)^{n-1} + q_{22}(x + 2a)^{n-2} + \cdots + q_{2n} \\
E^a f_1(x) &= (p_1 - q_{11})(x + 2a)^{n-1} + (p_2 - q_{12})(x + 2a)^{n-2} + \cdots + (p_n - q_{1n}) \\
\Delta_a f_1(x) &= (p_2 - q_{12} - q_{22})(x + 2a)^{n-2} + \cdots + (p_n - q_{1n} - q_{2n})
\end{aligned}$$

and so on for the higher orders, since $\Delta_a f_{s-1}(x) = \Delta_a^s f(x)$. In the practical application of this method, "a" may be conveniently taken as unity, and an abridged form of synthetic division employed. Thus, if

$$f(x) = 5x^4 + 3x^3 + 7x^2 - 2x + 3, \text{ then}$$

$$\begin{array}{rrrr:rl}
5 + & 3 + & 7 - & 2 & + \quad 3 & = f \\
 & -\ 2 + & 9 - & \mathbf{11} & + \ 14 & \\
 & -\ 7 + & 16 - & 27 & & \\
 & -\ 12 + & 28 & & & \\
 & -\ 17 & & & & \\
\hline
 & 20 & -\ 21 + & 25 & - \quad \mathbf{11} & = f_1 \\
 & & -\ 41 + & \mathbf{66} & - \quad 77 & \\
 & & -\ 61 + & 127 & & \\
 & & -\ 81 & & & \\
 & & 60 & -\ 102 & + \quad 66 & = f_2 \\
 & & & -\ \mathbf{162} & + \ 228 & \\
 & & & \underline{-\ \mathbf{222}} & & \\
 & & & \mathbf{120} & -\ \mathbf{162} & = f_3 \\
 & & & & \underline{-\ 282} & \\
 & & & & \mathbf{120} & = f_4.
\end{array}$$

As is evident from the darkened numerals, all figures to the right of the dotted line are redundant and may be omitted. From the above,

$$\Delta f(x) = 20(x+1)^3 - 21(x+1)^2 + 25(x+1) - 11$$

$$\Delta^2 f(x) = 60(x+2)^2 - 102(x+2) + 66$$

$$\Delta^3 f(x) = 120(x+3) - 162$$

$$\Delta^4 f(x) = 120.$$

SOME PRACTICAL INTERPOLATION FORMULAS

BY JOHN L. ROBERTS

Sometimes we wish to find by means of interpolation an approximation to a particular value of w_x in the interval between the known values, w_0 and w_1. But it also might be desirable in the interval from w_0 to w_1 to interpolate several approximations to w_x at equidistant values of x. It is very important to know that a formula which might be very satisfactory to interpolate a particular value in an interval might seriously fail to be the most satisfactory formula when it is desired to interpolate several values in the same interval. The range of this paper is so limited that we only wish to find by means of interpolation several approximations to the true value of w_x in the interval from w_0 to w_1 at equidistant values of x.

One way to perform an interpolation of this sort is to use osculatory interpolation.[1] The real function of osculatory interpolation is to secure smoothness at the known points, which are sometimes called pivotal points. By roughness is meant that one or more of the successive derivatives are discontinuous at the pivotal points. Experience proves that the osculatory formulas usually secure smoothness either at the expense of labor or by a loss of accuracies over the entire range from w_0 to w_1. Frequently the function of interpolation formulas is to save labor. In many cases it appears reasonable to save labor by a loss of both smoothness and accuracy. Formulas are herein selected, without direct regard for smoothness, so as to secure the best possible compromise between a maximum of accuracy and a minimum of labor. It appears that this results in many cases in a loss of smoothness that is no more objectionable than the loss in accuracy.

The actuarial profession, while trying to perfect their methods of constructing mortality tables, have made contributions of a high order of scholarship to the theory of osculatory interpolation. But since the statistician, the astronomer, the physicist, and other scientists also have occasions to make interpolations, it seems to be very important to discuss the problem of finding the most practical methods of interpolation, not only from the special viewpoint of the actuary, but also from the general viewpoint of mathematics.

Δw_x is called the first difference of w_x, and may be defined by $\Delta w_x = w_{x+1} - w_x$.

[1] Since this paper presupposes certain knowledge on the part of the reader, it may be worth while to indicate some sources of this knowledge. The elementary parts of this knowledge can be found in any good book on finite differences "Population Statistics and Their Compilation" by Hugh H. Wolfenden, published by the Actuarial Society of America, contains an excellent summary of osculatory interpolation. This summary indicates some valuable sources of information.

Second, third, and higher differences are merely successive differences of the first. When use is made of central difference interpolation formulas, it is convenient to adopt Woolhouse's notation, which is defined by means of the following equations: $\Delta w_{-2} = a_{-2}$, $\Delta w_{-1} = a_{-1}$, $\Delta w_0 = a_1$, $\Delta w_1 = a_2$, $\Delta^2 w_{-2} = b_{-1}$, $\Delta^2 w_{-1} = b_0$, $\Delta^2 w_0 = b_1$, $\Delta^3 w_{-2} = c_{-1}$, $\Delta^3 w_{-1} = c_1$, $\Delta^4 w_{-2} = d_0$, $\Delta^5 w_{-2} = e_1$, $\Delta^6 w_{-3} = f_0$, etc.

An important family of curves can be represented by

$$u_x = u_0 + xa_1 + \frac{1}{2}x(x-1)B + \frac{1}{6}x(x-1)\left(x-\frac{1}{2}\right)C. \qquad (1)$$

Assume $u_0 = w_0$ and $\Delta u_0 = \Delta w_0$. Then a study of (1) shows that a_1, which has already been defined, must be a factor in the second term in order that (1) may be satisfied when $x = 1$. (1) is a third degree equation. However, if $C = 0$, (1) becomes a second degree equation; if both $B = 0$ and $C = 0$, (1) becomes a first degree equation. In other words, by giving B and C proper values, (1) can be made to become many different interpolation formulas.

For many purposes interpolation by a first degree formula is not sufficiently accurate. . We, therefore, might wish to interpolate by either a second or a third degree formula. Since it is possible to draw an unlimited number of second degree curves or third degree curves between the points P_0 and P_1, the problem of selecting the best second degree interpolation curve and the best third degree curve is of great practical importance.

I

Suppose that w_{-2}, w_{-1}, w_0, w_1, w_2, and w_3 can be found in a table of values of the function w_x, and that we wish to find by means of interpolation several approximate values of w_x in the interval from w_0 to w_1. These six given values of w_x can be used to determine six pivotal points, which determine a fifth degree curve. Suppose this curve represents the function v_x. Then w_x and v_x would have exactly the same values at the six pivotal points, but would have values which are only approximately the same at other points. Using the first six terms of the Gauss central difference interpolation formula, we have

$$v_x = v_0 + xa_1 + \frac{1}{2!}x(x-1)b_0 + \frac{1}{3!}(x+1)x(x-1)c_1$$
$$+ \frac{1}{4!}(x+1)x(x-1)(x-2)d_0$$
$$+ \frac{1}{5!}(x+2)(x+1)x(x-1)(x-2)e_1.$$

It is proper to use in this formula the differences a_1, b_0, etc., which have already been defined as differences of w_x because these differences are exactly equal to the corresponding differences of v_x. Suppose P_0, $P_{\frac{1}{3}}$, $P_{\frac{2}{3}}$, and P_1 are four points

which are determined by v_x. Then B and C can be determined so that (1) will represent the curve which can go through these four points.
Then

$$u_{\frac{1}{3}} = u_0 + \frac{1}{3} a_1 - \frac{1}{9}\left(B - \frac{1}{18} C\right)$$

and

$$v_{\frac{1}{3}} = u_0 + \frac{1}{3} a_1 - \frac{1}{9}\left(b_0 + \frac{4}{9} c_1 - \frac{5}{27} d_0 - \frac{7}{81} e_1\right).$$

Also

$$u_{\frac{2}{3}} = u_0 + \frac{2}{3} a_1 - \frac{1}{9}\left(B + \frac{1}{18} C\right)$$

and

$$v_{\frac{2}{3}} = u_0 + \frac{2}{3} a_1 - \frac{1}{9}\left(b_0 + \frac{5}{9} c_1 - \frac{5}{27} d_0 - \frac{8}{81} e_1\right).$$

Since $u_{\frac{1}{3}} = v_{\frac{1}{3}}$ and $u_{\frac{2}{3}} = v_{\frac{2}{3}}$, we have two equations, which can be solved for B and C.

$$B = b - \frac{5}{27} d \text{ and } C = c_1 - \frac{1}{9} e_1 \tag{2}$$

where b and d are defined by

$$b = \frac{1}{2}(b_0 + b_1) \text{ and } d = \frac{1}{2}(d_0 + d_1).$$

A study of (1) shows that $u_{\frac{1}{2}}$ does not depend upon C because the term containing C becomes zero when $x = \frac{1}{2}$, and also shows that u_x over the entire range from u_0 to u_1 is more sensitive to errors in B than errors in C. The B in (2) usually contains some error because the six terms of the Gauss formula which were used in determining B usually produce results which are only approximate. Consequently a comparatively large error in C would not produce an important error.
Assume

$$B = b - \frac{5}{27} d \text{ and } C = c_1 - \frac{5}{27} e_1. \tag{3}$$

B is the same in both (2) and (3), but C is not the same. The accuracy of (2) and the accuracy of (3) do not differ by an important amount. On the other hand, if any attempt to apply (2) is compared with the working illustrations of (3) in this article, it will be found that (2) to an important extent is more laborious than (3). Therefore (3) is a better compromise between a maximum of accuracy and a minimum of labor than (2). For this reason (2)

ought not to be regarded as a practical formula. On the other hand (2) because of its great accuracy serves as an ideal with which other formulas can be compared. In other words (2) is of theoretical importance.

In like manner another interpolation formula can be found if we use the first four terms of the Gauss formula to determine $P_{\frac{1}{2}}$. Then

$$u_{\frac{1}{2}} = u_0 + \frac{1}{2}a_1 - \frac{1}{8}B$$

and

$$v_{\frac{1}{2}} = u_0 + \frac{1}{2}a_1 - \frac{1}{8}\left(b_0 + \frac{1}{2}c_1\right).$$

Since $u_{\frac{1}{2}} = v_{\frac{1}{2}}$, we can solve for B, and C is left arbitrary. If $C = 0$, we again get an excellent compromise between a maximum of accuracy and a minimum of labor. The following second degree formula results.

$$B = b \text{ and } C = 0. \tag{4}$$

In order that the value of (3) and (4) may be appreciated, they are herein compared with some other formulas which have been of historical importance.

If the point $P_{\frac{1}{2}}$ can first be accurately determined, a second degree curve through the points P_0, $P_{\frac{1}{2}}$, and P_1 would probably give more accurate results than such a curve through the points P_0, P_1, and P_2 because the first three points are in a smaller neighborhood; the second curve can be represented by the first three terms of the Gregory-Newton interpolation formula. The points P_{-1}, P_0, P_1, and P_2 determine a third degree curve, which can be represented by the first four terms of the Gauss central difference formula. It is probable that these terms would determine $P_{\frac{1}{2}}$ much more accurately than the first three terms of the Gregory-Newton formula because the latter is not a central difference formula with respect to $P_{\frac{1}{2}}$ and because four terms usually give more accurate results than only three terms. Consequently there is a strong probability that (4) is more accurate than the first three terms of the Gregory-Newton formula. In like manner (4) is more accurate than the first three terms of the Gauss formula. It is interesting to observe that (4) is the first three terms of the Newton-Bessel formula.

$$\text{If } B = b \text{ and } C = 3c_1,$$

then (1) is equivalent to Karup's osculatory interpolation formula in terms of differences taken centrally. B is the same in both (4) and Karup's formula. No interpolation formula can be very accurate unless C is about equal to c_1. Since, then, the error in C in Karup's formula is about twice as great as the error in C in (4), his formula is distinctly less accurate than (4). Since (4) is a second degree curve and Karup's formula is a third degree curve, his formula is very much more laborious. (4) is extremely accurate for a formula having its labor saving properties; for many purposes its roughness and inaccuracy appear to

be in about the right proportion. On the other hand Karup's formula is extremely inaccurate for a formula so laborious; its only good point is its smoothness.

Changing somewhat the meanings of u and w, (3) may be written

$$
\begin{aligned}
u_{x+n} &= u_n + x\Delta u_n \\
&+ \frac{1}{2}x(x-1)\left[\frac{1}{2}(\Delta^2 w_n + \Delta^2 w_{n-1}) - \frac{5}{54}(\Delta^4 w_{n-1} + \Delta^4 w_{n-2})\right] \\
&+ \frac{1}{6}x(x-1)\left(x-\frac{1}{2}\right)\left(\Delta^3 w_{n-1} - \frac{5}{27}\Delta^5 w_{n-2}\right).
\end{aligned}
$$

If

$$\frac{du}{dx} = u'_{x+n},$$

then

$$u'_{0+0} - u'_{1-1} = \frac{1}{54}d_0 + \frac{5}{162}f_0,$$

which is the amount of discontinuity in $\frac{du}{dx}$ at P_0. (3) has greater smoothness than (4); in other words (3) is more like an osculatory formula. On the other hand

$$B = b - \frac{1}{6}d \text{ and } C = c_1 - \frac{1}{6}e_1, \tag{5}$$

which is equivalent to an important osculatory interpolation formula by Mr. Robert Henderson, compares much better with (3) from the viewpoint of labor saving and accuracy than Karup's formula does with (4).

II

An excellent formula can be easily spoiled if the method of applying it is not practical. Mr. Henderson, in the Transactions of the Actuarial Society of America, Vol. IX, applies (5) in such a way that the numerical work is very convenient. Some writers seem to have been very careless about this matter. A method intended to interpolate several values between w_0 and w_1 should provide that the end value w_1 shall be exactly reproduced if no error is made in the computation. In other words a good method should provide a check upon the work. At the same time, in order to avoid unnecessary labor, the work should not retain unnecessary decimal places or figures. In other words fictitious accuracy should be avoided. The following working illustrations are intended to show good methods of application of formulas and to show how much labor is necessary in order to apply them; also the size of the errors can be used to illustrate the theory.

When (4) is applied at either end of the table, where terms are not available for the calculation of the differences required, it should be assumed that the fourth differences that cannot be computed vanish and the required differences should be filled in consistently with that assumption. Δw_x represents the first differences. But it is convenient to have S represent the first differences in such a manner that they are arranged centrally in the working illustration. S^2 in like manner represents the second differences. The 2 in S^2 means S^2 is a second difference, and does not have the familiar meaning used in algebra. In the case of (4), $\Delta u_x = a_1 + xB$, $\Delta^2 u_x = B$, and the higher differences all equal zero. Since we wish in the working illustration of (4) to interpolate four values between w_0 and w_1, δ and δ^2 are defined by $\delta u_x = u_{x+.2} - u_x$ and $\delta^2 u_x = \delta u_{x+.2} - \delta u_x$. It is proved in any good book on finite differences that there are possibilities that Δ and δ, which are symbols of operation, can be separated from the functions upon which they operate, and they can be treated as if they were algebraic numbers. Consequently $1 + \delta = (1 + \Delta)^{\frac{1}{5}}$. In other words by means of the binomial law $\delta u_x = (.2\Delta - .08\Delta^2)u_x$, where all the terms within the parenthesis are to be considered as operating [illegible] $\delta^2 u_x = .04\Delta^2 u_x$. s, s_x, and s^2 are defined by $s = s_x = \delta u_x$, an[illegible]erefore the middle $s = \delta u_{.4} = .2a_1$, and $s^2 = .04B = .02(b_0 + b_1)$. We are now in position to apply (4) to the case when $w_x = (1.04)^n$. It might prevent confusion if it is stated that x and n are related to each other in such a way that we always interpolate between w_0 and w_1.

n	$(1.04)^n$	s	S	S^2	s^2
80	23.050	.9218		.845	
81	23.9718	.9603			
82	24.9321	.9988	4.994		.0385
83	25.9309	1.0373			
84	26.9682	1.0758			
85	28.044	1.1190		1.081	
86	29.1630	1.1670			
87	30.3300	1.2150	6.075		.0480
88	31.5450	1.2630			
89	32.8080	1.3110			
90	34.119	1.3636		1.317	
91	35.4826	1.4210			
92	36.9036	1.4784	7.392		.0574
93	38.3820	1.5358			
94	39.9178	1.5932			
95	41.511			1.553	

Some of the explanation of the application of (4) applies to (3) and does not need to be repeated. The method herein used of applying (3) is either the same as or a development of the Henderson method of applying (5). If it is desired to apply (3) at either end of the table, where terms are not available for the calculation of the differences required, it can be assumed that the sixth differences that can not be computed vanish and the required differences can be filled in consistently with that assumption. A study of the theory underlying this assumption shows that it does not result in a true central difference formula and that it consequently results usually in some loss of accuracy. In the case of (3) before the finding of the differences of (1), it is convenient to write it as follows:

$$u_x = u_0 + xa_1 + \frac{1}{2}x(x-1)\left(B + \frac{1}{2}C\right) + \frac{1}{6}x(x-1)(x-2)C.$$

Then

$$\Delta u_x = a_1 + x\left(B + \frac{1}{2}C\right) + \frac{1}{2}x(x-1)C,$$

$$\Delta^2 u_x = \left(B + \frac{1}{2}C\right) + xC, \text{ and } \Delta^3 u_x = C.$$

Suppose we wish to interpolate four values between w_0 and w_1. δ and δ^2 have already been defined. $\delta^3 u_x = \delta^2 u_{x+2} - \delta^2 u_x$. Then $1 + \delta = (1 + \Delta)^{\frac{1}{5}}$, or $\delta u_x = (.2\Delta - .08\Delta^2 + .048\Delta^3)u_x$. Also $\delta^2 u_x = (.04\Delta^2 - .032\Delta^3)u_x$ and $\delta^3 u_x = .008\Delta^3$. s^2, s_x^2, and s^3 are defined by $s^2 = s_x^2 = \delta^2 u_{x-2}$, and $s^3 = s_4^3 = \delta^3 u_x$. The first

$$s^2 = \delta^2 u_{-.2} = .04\left(B - \frac{1}{2}C\right) = .04\left(b_0 - \frac{5}{27}d_0\right).$$

The last

$$s^2 = \delta^2 u_{.8} = .04\left(B + \frac{1}{2}C\right) = .04\left(b_1 - \frac{5}{27}d_1\right).$$

.1852 might be a useful approximation to $\frac{5}{27}$. The remaining s^2's, s should be filled in so that they are in arithmetical progression with irregularities at the ends. If the irregularities can be distributed equally at both ends, the irregularities cause an error in C, but none in B. Errors in B are more important than those in C. The middle $s = \delta u_{.4} = .2a_1 - s^3$. In the following working illustration, $w_x = \sin n$.

In the sixth line from bottom

n	$\sin n$	S	S^2	S^3	S^4
−60	−.86603				
		.36603			
−30	−.50000		.13397		
		.50000		−.13397	
0	.00000		.00000		.00000
		.50000		−.13397	
30	.50000		−.13397		.03588
		.36603		− 09809	
60	.86603		−.23206		
		.13397			
90	1.00000				

n	$\sin n$	s	s^2	
0	.00000	.104498	.000000	
6	.104498	.103374	− 001124	
12	.207872	.101125	2249	−.001125
18	308997	.097751	3374	
24	.406748	93252	4499	
30	.50000		−.005624	

Suppose we wish to interpolate nine values between w_0 and w_1 by the use of (3). Then $\delta u_x = u_{x+.1} - u_x$, $\delta^2 u_x = \delta u_{x+.1} - \delta u_x$, and $\delta^3 u_x = \delta^2 u_{x+.1} - \delta^2 u_x$. Consequently $1 + \delta = (1 + \Delta)^{1/10}$, or $\delta u_x = (.1\Delta - .045\Delta^2 + .0285\Delta^3)u_x$. Then $\delta^2 u_x = (.01\Delta^2 - .009\Delta^3)u_x$ and $\delta^3 u_x = .001\Delta^3$. $s^2 = s_x^2 = \delta^2 u_{x-.1}$ and $s^3 = s_4^3 = \delta^3 u_x$. The first

$$s^2 = \delta^2 u_{-.1} = .01\left(B - \frac{1}{2}C\right) = .01\left(b_0 - \frac{5}{27}d_0\right).$$

The last

$$s^2 = \delta^2 u_{.9} = .01\left(B + \frac{1}{2}C\right) = .01\left(b_1 - \frac{5}{27}d_1\right).$$

$\delta u_{.4} = (.1a_1 - 4s^3) - \frac{1}{2}\delta^2 u_{.4}$ and $\delta u_{.5} = (.1a_1 - 4s^3) + \frac{1}{2}\delta^2 u_{.4}$.

	sin n	s	s^2	s^3
0	.00000	52318	.000000	
3	.052318	52179	−.000139	
6	.104497	51899	280	
9	.156396	51478	421	
12	.207874	.050916	562	−.000141
15	.258790	.050212	703	
18	.309002	49368	844	
21	.358370	48383	985	
24	406753	47257	1126	
27	.454010	45990	1267	
30	.50000		−.001406	

Suppose we wish to interpolate five values between w_0 and w_1. The first $s^2 = \frac{1}{36}\left(b_0 - \frac{5}{27}d_0\right)$ and the last $s^2 = \frac{1}{36}\left(b_1 - \frac{5}{27}d_1\right)$.

$$\delta u_{\frac{1}{2}} = \frac{1}{6}(a_1 - 8\delta^3 u_x) - \frac{1}{2}\delta^2 u_{\frac{1}{2}}.$$

and

$$\delta u_{\frac{1}{2}} = \frac{1}{6}(a_1 - 8\delta^3 u_x) + \frac{1}{2}\delta^2 u_{\frac{1}{2}}.$$

In the following working illustration the given values of sin n are written correct to five decimal places; in other words after each decimal point there are five symbols or digits representing numbers; also each of these symbols is written in the scale of ten. It can be observed that some values of u_x, s, s^2, and s^3 in the working illustration have six symbols to the right of the decimal point, and that some values have seven symbols to the right of the decimal point. In all cases the sixth symbol to the right of the decimal point is written in the scale of ten, and the seventh symbol is written in the scale of six. This procedure provides a check by exactly reproducing w_1. Also this procedure does not cause much fictitious accuracy, and can be quickly used after a little practice.

n	sin n	s		
0	.00000	87130	.000000	
5	.0871305	86479	−.000651	
10	.1736104	.0851775	1302	
15	.2587883	.0832245	1953	−.000651
20	.3420132	80620	2604	
25	.4226341	77365	3255	
30	.50000		.003906	

In general if we wish to interpolate $i - 1$ values between w_0 and w_1 when i is neither five nor ten, w_1 can be exactly reproduced if some of the symbols are written in the scale of i. If $i = 12$, it is evident that we need two extra symbols, say t and e, to stand for ten and eleven respectively. If we wish to interpolate $i - 1$ values between w_0 and w_1 by the use of (4), in the computation each of u_x, s and s^2 except the given values should contain one more symbol than each given value contains, and the extra symbol should be written in the scale of i.

ON EVALUATING A COEFFICIENT OF PARTIAL CORRELATION

By Grace Strecker

It is to be shown here that when the multiple correlation coefficient $R_{n;12\cdots(n-1)}$ is found by the method of Horst[1] the partial correlation coefficient $R_{n(n-1);12\cdots(n-2)}$ can be found in terms of the β's. If we are interested only in evaluating a partial correlation between two variables, we may also employ the method which will be given here.

Without loss of generality the dependent variables may be chosen to be the nth and $(n-1)$st. The coefficient of partial correlation as given by Rietz[2] may be expressed in the following form:

$$R_{n(n-1);12\cdots(n-2)} = \sqrt{\frac{\dfrac{R_{(n-1)(n-1)}}{R_{(n-1)(n-1)nn}} - \dfrac{R}{R_{nn}}}{\dfrac{R_{(n-1)(n-1)}}{R_{(n-1)(n-1)nn}}}}. \tag{1}$$

$R_{(n-1)(n-1)}$ may be treated as a new determinant R'. Regarding its elements as the coefficients of a set of normal equations ($n - 1$ in all) whose constant terms are zero, we may follow through the Doolittle elimination process. For the case where $n = 4$ we have the table given below.

In comparing this outline with the one illustrating the Doolittle elimination process for R when $n = 4$ we see that

$$\gamma'_{11} = \gamma_{11} = \frac{A_{11}}{R^2},$$

$$\gamma'_{22} = \gamma_{22} = \frac{rA_{1122}}{R^2A_{11}},$$

$$\gamma'_{33} = \alpha'_{33} - \sum_2^3 \beta'_{i3} = \alpha_{44} - \sum_2^3 \beta_{i4}.$$

Therefore, we have

$$R' = \frac{A_{11}}{R^2} \cdot \frac{rA_{1122}}{R^2A_{11}} \cdot \left(\alpha_{44} - \sum_2^{3} \beta_{i4}\right)$$

$$= \prod^2 \gamma_{ii}\left(\alpha_{44} - \sum^3 \beta_{i4}\right).$$

[1] Horst Paul, *A Short Method for Solving for a Coefficient of Multiple Correlation*, Annals of Mathematical Statistics, Vol. III, No. 1, Feb. 1932, pp. 40–44.

[2] Rietz, H. L., *Mathematical Statistics*, p. 101.

Reciprocal	1	2	3	α	β	γ	δ
$-\frac{R^2}{A_{11}}$	$\frac{A_{11}}{R^2}$	$\frac{A_{12}}{R^2}$	$\frac{A_{14}}{R^2}$	α_1'		γ_1'	
	-1	$-\frac{A_{12}}{A_{11}}$	$-\frac{A_{14}}{A_{11}}$				δ_1'
		$\frac{A_{22}}{R^2}$	$\frac{A_{24}}{R^2}$	α_2'			
		$-\frac{A_{12}^2}{R^2A_{11}}$	$-\frac{A_{12}A_{14}}{R^2A_{11}}$		β_{22}'		
$-\frac{R^2A_{11}}{AA_{1122}}$		$\frac{AA_{1122}}{R^2A_{11}}$	$\frac{AA_{1124}}{R^2A_{11}}$			γ_2'	
		-1	$-\frac{A_{1124}}{A_{1122}}$				δ_2'
			$\frac{A_{44}}{R^2}$	α_3'			
			$-\frac{A_{14}^2}{R^2A_{11}}$		β_{23}'		
			$-\frac{AA_{1124}^2}{R^2A_{11}A_{1122}}$		β_{33}'		
			$\alpha_3' - \sum_{2}^{3} \beta_{i3}'$			γ_3'	
			-1				δ_3'

In the general case:

$$\gamma_{11} = \gamma_{11},$$

$$\gamma_{22}' = \gamma_{22},$$

$$\gamma_{(n-2)(n-2)} = \gamma_{(n-2)(n-2)},$$

$$\gamma_{(n-1)(n-1)}' = \alpha_{nn} - \sum^{n-1} \beta_{in}.$$

Hence

$$R_{(n-1)(n-1)} = R' = \prod_{1}^{n-2} \gamma_{ii}\left(\alpha_{nn} - \sum_{2}^{n-1} \beta_{in}\right).$$

Since $R = \prod^{n} \gamma_{ii}$, then $R_{(n-1)(n-1)nn} = \prod^{n-2} \gamma_{ii}$, from which we see that

$$\frac{R_{(n-1)(n-1)}}{R_{(n-1)(n-1)nn}} = \frac{\prod_{1}^{n-2} \gamma_{ii}\left(\alpha_{nn} - \sum_{2}^{n-1} \beta_{in}\right)}{\prod_{1}^{n-2} \gamma_{ii}} = \alpha_{nn} - \sum^{n-1} \beta_{in}.$$

But since $\alpha_{nn} = 1$, then

$$\frac{R_{(n-1)(n-1)}}{R_{(n-1)(n-1)nn}} = 1 - \sum_{2}^{n-1} \beta_{in}.$$

It has been shown that

$$\frac{R}{R_{nn}} = 1 - \sum_{2}^{n} \beta_{in}.$$

Substituting the above values for $\frac{R_{(n-1)(n-1)}}{R_{(n-1)(n-1)nn}}$ and $\frac{R}{R_{nn}}$ in equation (1), we have

$$R_{n(n-1),12\cdots(n-2)} = \sqrt{\frac{1 - \sum_{2}^{n-1} \beta_{in} - \left(1 - \sum_{2}^{n} \beta_{in}\right)}{1 - \sum \beta_{in}}}$$

or

$$R_{n(n-1);12\cdots(n-2)} = \sqrt{\frac{\beta_{nn}}{1 - \sum^{n-1} \beta_{in}}}.$$

Hence it is seen that when the β's given by Horst (page 42) are calculated, it is an easy matter to solve for the partial correlation $R_{n(n-1),12\cdots(n-2)}$.

ST. LOUIS UNIVERSITY.

A THEORY OF VALIDATION FOR DERIVATIVE SPECIFICATIONS AND CHECK LISTS[1]

By Lee Byrne

Visiting Professor of Secondary Education, New York University

Part I. Research Products Which May Be Classified as Derivative Specifications and Check Lists

Meaning of Specification

In specification something is assigned a specific character. The something to be thus assigned a specific character may be called the specificandum. The specific character assigned to the specificandum, or (as a second meaning) the act of so doing, may be called the specification.

A proposition is the smallest unit in which it is possible to embody a complete thought and is ordinarily represented by a single sentence. In specification the characterization may be confined to a single proposition or it may be extended to include an indefinitely large number of propositions. So a specification may be embodied in a sentence, a paragraph, a chapter, or a whole book. No matter how far it is extended it will never give complete determination, as our knowledge cannot be made exhaustive or our control be given an absolute precision.

In view of the meaning assigned to specification it is evident that very many books and monographs could in this sense be classified as specifications.

Meaning of Derivative Specification

There is a type of specification (book or monograph) which is developed by deriving it from a group or class of specifications which already exist. This class may be a total class of all such specifications, or a group of those accepted as authoritative, or a group of those taken to be representative. A specification derived in this manner may be called a derivative specification. As an example we could take almost any first-class work by a present-day historian; by historians it would be called "secondary" because it is based on study of pre-existent documents called "primary sources."

Meaning of Check List

The act of *deriving* a product from a pre-existent set of documents may, as we have seen, take the form of a derivative specification, embracing an as-

[1] This paper is an amplification of a report made in the statistical section of the American Educational Research Association at its meeting in February, 1931.

semblage of determinates or determinations. On the other hand the product derived may be intended merely to indicate the ground covered or to be covered by determination, without actually selecting the particular determinations. Such a product will be called a check list. The term is not a very happy one, but it is in very common use. If we think of a specification as an assemblage of determinations then a check list could be thought of as a corresponding set of determinables.[2] Since any determinable is capable of an indefinite number of determinations it is evident that a long check list could give rise to an extremely large number of different specifications, of which, of course, some fraction might prove undesirable, inadmissible, or false.

Modes of Specification: How We Specify

If we examine any specification to see how the specifying is done we shall find that it ultimately takes the form of specification under aspects. The following diagram indicates the principal (perhaps all the) possibilities in the way of specification.

Naming the original or main specificandum
- Naming an aspect
  - Characterization of the specificandum under the aspect named
- Naming a relation (includes process, operation etc.)
  - Naming an aspect of the relation
    - Characterization of the relation under aspect named
  - Naming a relatum or thing related (a new specificandum)
    - Naming an aspect of the relatum
      - Characterization of relatum under aspect named
- Naming a part
  - Naming an aspect of the part
    - Characterization of the part under aspect named

(The naming of aspects may be merely implicit but it is always present in principle.)

---

[2] On the notion of the "determinable," which is due to W. E. Johnson, see his Logic, Cambridge University Press (1921), Part I, p. xxxv and Chapter XI.

Thus it appears that if specification is pressed far enough it always ultimately becomes specification under aspects. Aspect and determinable may be regarded as synonyms.

Current Examples of Derivative Specifications and Check Lists

At the present time it will be found that we have very many products of research which take forms capable of being classified as some kind of derivative specification or (derivative) check list in the senses in which these expressions have been explained.

I have distinguished more than twenty different logical types of derivative specification or check list which are exemplified in the current literature of educational research and related subjects. However space will not permit exhibition of examples of these different types.

Part II. Validation of Derivative Specifications and Check Lists

Many research products may be classified as derivative specifications or check lists, derivative in the sense that they have been derived from a group of documents (books, articles, journals, newspapers, courses of study, etc.) through analysis of their content. Such source documents themselves we shall call specifications or groups of specifications.

The only validation problem raised here is the question whether the resulting check list or derivative specification truly represents the class of source specifications used. The further question whether the class of source specifications itself constitutes a satisfactory source is not discussed.

From this point of view, if a check list or derivative specification is based in some suitable manner on *all* the documents of the class represented, no real validation problem arises; the validity has to be regarded as perfect.

It may often happen that the investigator does not wish to analyse *all* of the specifications of the class in question but prefers to save time and labor by confining his analysis to a select group drawn from the total class as a sample. In this case the problem arises as to how far results based on such sample should be judged to be truly representative of the entire class of specifications (most of which have not been analysed). A problem of this nature may be called the problem of validity for this kind of work.

Such a validation problem appears to take the same form whether the product to be validated is a derivative specification or (derivative) check list. Accordingly we shall for the sake of brevity carry on the discussion by referring to the problem as that of validating (derivative) check lists. The same principles would apply if the product happened to be a derivative specification.

In order to consider the validity of a check list based on a sample group of specifications (called here a Sample Check List) we may hypothesize a check list based in the same manner on the entire class of specifications from which the sample was drawn. Such a hypothetical check list (which is not made) will be called the Ideal Check List. Then the problem of validity may be con-

ceived as the question as to how far the content of the Sample Check List agrees with the unknown content of the Ideal Check List.

An overlapping of the two appears ordinarily to be certain but a failure of complete coincidence is very highly probable. The question is what degree of coincidence is to be expected.

This general validity problem naturally divides into two separate questions. The first question asks what proportion of the content of the Sample Check List may be expected to be present also in the Ideal Check List; this may be called the (sub-) problem of reliability. The second question asks what proportion of the content of the Ideal Check List may be expected to be present in the Sample Check List; this may be called the (sub-) problem of completeness. The answers to these two problems, if expressed in numerical percentages, could be called the Index of Reliability and Index of Completeness respectively.

We shall first consider these two problems in their simplest form and afterward in a more complex form in which they exhibited themselves in a recent study by the writer.[3] The simple case presents no great difficulty and it is possible that a different method of disposing of it might be preferred. The more complex case, however, appears to be rather difficult of solution and the writer has not been able to find in the literature any developed technique for handling it. The simple case is presented here primarily because it affords, by further extension, a successful approach to the difficult problem of the more complex case.

Simple Case

Terms and Symbols

The "class of specifications" will be understood to consist of all specifications which belong to the whole class of specifications regarded as a source, a class which we claim to represent in our final product. In this problem the "class" will not be regarded as indefinitely large but as consisting of a definite number of specifications, a number to be ascertained by actual count or by careful estimate.

"Sample specifications" are the limited group selected from the class for purposes of actual analysis, and which play the rôle of representing the whole class. The remaining specifications of the class are not analyzed.

"Sample Check List Material" is a name for the assemblage of all the different items found in one or more sample specifications.

"Ideal Check List Material" is a name for a hypothetical assemblage of all the different items found in one or more specifications in the class. Only those appearing in some sample specifications can be actually known, the rest are hypothetical.

[3] Byrne, L. Check List Materials for Public School Building Specifications. Teachers College, Columbia University. 1931.

Write

M (constant) = total number of specifications in class
N (variable) = number of these specifications in which a particular item under consideration appears (this number is hypothetical and some of the particular items themselves are hypothetical)
m (constant) = number of sample specifications
n (variable) = number of sample specifications in which a particular (the same) item appears

Values of n may be expected to vary for different items, from m to 0 by intervals of 1, the zero value appertaining to any item wholly absent from the Sample Check List Material (hypothetically present in Ideal Check List Material).

Values of N might be expected to vary, for different items, from M to 1 by intervals of 1. But in this problem the convention will be adopted that the range is from M downward by intervals of $\frac{M}{m}$. Thus if the number M should be five times as large as the number m then the range for N would be treated as proceeding from M downward by intervals of 5: M, $M - 5$, $M - 10$, $\cdots$ 5.

A "tabulation" will mean a statistical table showing how many different items appear in every possible number of specifications. A tabulation must be made by actual count for the items of the sample specifications, and will show the number of items having each possible value of n. A similar tabulation is hypothetical for the items in all the specifications of the class, that is for the number of items having each value of N permitted by the convention of the last paragraph.

"Tabulation cell" (or simply "cell") will mean, as needed, either the number of items or the group of items appearing in any designated number of specifications. For Sample Check List Material it will be the number or group of items to which a particular value of n appertains; for Ideal Check List similarly the number of items or group of items to which a particular value of N appertains (hypothetically).

"Sample Check List" will mean a list of items selected from the Sample Check List Material according to some adopted criterion. For illustrative purposes we shall consider this criterion to be, for example, the numerical ratio $n \geqq \frac{m}{2}$.

"Ideal Check List" will mean a list of items selected from the Ideal Check List Material according to some adopted criterion. For illustrative purposes we shall consider this criterion to be the numerical ratio $N \geqq \frac{M}{2}$.

Problem of Reliability

The problem of reliability may be restated and renamed the General Reliability Problem. This may be broken up into a group of problems which will

be called Elementary Reliability Problems. Each of the latter may be in turn broken up into a group of problems which will be called Ultimate Reliability Problems. Each Ultimate Reliability Problem may be solved directly. Combination of these solutions will yield solutions of the Elementary Reliability Problems. Combinations of the latter solutions will finally yield the solution of the General Reliability Problem.

These problems will now be stated

General Reliability Problem: What proportion of the items present in Sample Check List may be expected to be present also in Ideal Check List?

Elementary Reliability Problem: What proportion of the items in a particular cell in Sample Check List may be expected to be present also in Ideal Check List?

Ultimate Reliability Problem: What proportion of the items in a particular cell in Sample Check List may be expected to be present also in some designated cell in Ideal Check List?

To solve an Ultimate Problem:

From the Fundamental Theorem in the Theory of Inductive Probability (Whittaker, E. T. and Robinson, G. The Calculus of Observations. London: Blackie & Son. 1924. p. 305) the solution may be expressed as

$$\frac{P_R \cdot p_s}{\Sigma Pp}.$$

Whittaker and Robinson's statement of the Fundamental Theorem in the Theory of Inductive Probability is as follows (form slightly changed without change in meaning):

"Suppose that a certain observed phenomenon may be accounted for by any one of a certain number of hypotheses, of which one, and not more than one, must be true: suppose moreover that the probability of the R-th hypothesis, as based on information in our possession before the phenomenon is observed, is P_R, while the probability of the observed phenomenon, on the assumption of the truth of the R-th hypothesis, is p_s. Then when the observation of the phenomenon is taken into consideration, the probability of the R-th hypothesis is

$$\frac{P_R \cdot p_s}{\Sigma Pp}$$

where the symbol Σ denotes the summation over all the hypotheses."[4]

It is clear that an Ultimate Reliability Problem is a case falling under this Fundamental Theorem. The observed phenomenon is any item occurring in any specified cell of Sample Check List, say cell $n = s$. It may be accounted for by a certain number of hypotheses as to its source in the Ideal Check List

[4] For the fundamental position of this theorem in a theory of science and for its proof one may also consult Jeffreys, H. Scientific Inference. Cambridge: Cambridge University Press. 1931. Chapter II (section 2.34).

Material; the different cells in the Ideal Check List Material are these different hypotheses of origin, hypothetical because we do not *know* from which one it has come but only that it must have come from some one of them; the cell from which it actually comes is the true hypothesis, though we do not know which one that is. That the origin of the item is in cell $N = R$ is the R-th hypothesis, and its probability is written P_R. The probability of the occurrence of the phenomenon on the assumption of the truth of the R-th hypothesis is the probability that an item in cell $N = R$ will appear in Sample Check List in cell $n = s$ and its probability is written p_s. As we clearly have in our Ultimate Reliability Problem a case falling under the Fundamental Theorem quoted we may accept as the required solution of the Ultimate Reliability Problem the formula already given in the initial statement:

$$\frac{P_R \cdot p_s}{\Sigma Pp}.$$

This expresses the probability that any item found in Sample-Check-List cell $n = s$ comes from (and appears in) Ideal-Check-List-Material cell $N = R$, or it gives the proportion of items found in Sample-Check-List cell $n = s$ that may be expected to come from (or appear in) Ideal-Check-List-Material cell $N = R$.

Meaning of any value of P (say P_R) = the probability that any item, drawn at random from those cells of Ideal Check List Material which are possible sources of items in Sample-Check-List cell $n = s$, will happen to be drawn from cell $N = R$.

Meaning of any value of p (say p_s) = the probability that any item in Ideal-Check-List cell $N = R$ will also be present in Sample-Check-List cell $n = s$. (Important: this supposition is *not* equivalent to its converse.)

Evaluation of P_R:

$$P_R = \frac{\text{number of items in cell } N = R}{\text{number of items in all cells which are possible sources of items in cell } n = s}.$$

For this ratio it is necessary to assume that the shape of the numerical curve formed by the group of Ideal-Check-List-Material cells is the same as that of the numerical curve formed by the group of Sample-Check-List-Material cells. On this assumption we may replace the numerator by the number of items in the Sample-Check-List-Material cell having an abscissa corresponding to that of the Ideal-Check-List-Material cell $N = R$, and replace the denominator by the sum of the numbers of items in all the cells with abscissae corresponding to those of Ideal-Check-List-Material cells which are possible sources of items in cell $n = s$.

Evaluation of p_s:

By the aid of "the definition of probability which is used in practically all treatises on the subject" (Coolidge, J. L. An Introduction to Mathematical

Probability. Oxford: Oxford University Press. 1925. p. 4) and the principle underlying the Theory of Combinations (Whitworth, W. A. Choice and Chance. New York: G. E. Stechert & Co. 1927. Proposition II) we are able to arrive at the evaluation:

$$p = \frac{C_{m-n}^{M-N} C_n^N}{C_m^M}$$

in which, for any p (say p_s), we employ for N the value $N = R$, and for n the value $n = s$. As the denominator later cancels out it may be disregarded throughout, simplifying the formula to

$$p = C_{m-n}^{M-N} C_n^N.$$

(A symbol such as C_n^N is read "the number of combinations of N things taken n at a time"; also written in several other forms.)

The definition referred to may be worded as follows (Coolidge's own preferred definition is not quite the same):

"An event can happen in a certain number of ways, which are all equally likely. A certain proportion of these are classed as *favorable*. The ratio of the number of favorable ways to the total number is called the probability that the event will turn out favorably."

The principle underlying the Theory of Combinations may be quoted from Whitworth as follows (also found in ordinary works on algebra):

"If one operation can be performed in m ways, and then a second can be performed in n ways, and then a third in r ways, (and so on), the number of ways of performing all the operations will be $m \times n \times r \times$ etc."

If it is not at once clear that the formula for evaluation of p follows from the definition and principle just quoted, the following considerations should make it evident.

We are working in terms of a particular item belonging to a particular Ideal-Check-List-Material cell, say cell $N = R$. "Favorable" occurrence requires that this item fall in a particular Sample-Check-List cell, say $n = s$, while falling in any other Sample-Check-List-Material cell (including cell $n = 0$ for absence) is "unfavorable." Again the real meaning of the "favorable" occurrence is that the item will be found in just $n = s$ out of the m specifications of the sample, and absent in the remaining $m - n$ specifications of the sample. Moreover presence in Ideal-Check-List-Material cell $N = R$ means that the item occurs in just $N = R$ of the M specifications that constitute the whole class and is absent in $M - N$ of these specifications. The total number of all the ways (favorable and unfavorable) in which our event can happen means the same as the total number of all the ways in which a group of m specifications can be selected from a larger group of M, and this is, of course, written C_m^M and given us in our denominator. The number of favorable ways in which our event can happen means the same as the number of ways in which N specifications containing the item can form groups of n specifications while at the

same time $M - N$ specifications not containing the item can form groups of $m - n$ specifications; the first distribution can be done in C^N_n ways and the second in C^{M-N}_{m-n} ways, so by Whitworth's principle the number of ways which these things can happen simultaneously is $C^{M-N}_{m-n}\, C^N_n$. Assembling numerator and denominator we have the formula initially stated for evaluation of p, viz.:

$$p = \frac{C^{M-N}_{m-n}\, C^N_n}{C^M_m}.$$

This is the general formula; in applying to the particular example $N = R$, $n = s$ the replacements for N and n, of course, give

$$p_s = \frac{C^{M-R}_{m-s}\, C^R_s}{C^M_m}.$$

Having a means of evaluating P and p we may solve all needed Ultimate Problems. The resulting solutions of the needed Ultimate Reliability Problems (not necessarily completed) enables us to arrive at the solution of any needed Elementary Reliability Problem in the form of a percentage which may be called an Index of Reliability for the Sample-Check-List cell in question. In computing this percentage we distinguish source-cells that belong to the Ideal Check List from other source-cells that belong to the Ideal Check List Material but not to the Ideal Check List.

By properly averaging cell-Indices of Reliability (which are really Indices of Reliability for the individual items in the cells) we may obtain a solution of the General Problem of Reliability in the form of an Average Index of Reliability for the Sample Check List as a whole.

In addition to the Average Index of Reliability for *the* Sample Check List we may easily secure also Average Indices of Reliability for any series of briefer Sample Check Lists selected from *the* Sample Check List, by properly averaging the Indices of cells contained in any Sample Check List in question, keeping the original criterion for Ideal Check List.

In practice it may not be necessary to compute all cell-Indices, as a portion of these may be entered in tables by any methods of interpolation regarded as acceptable.

Problem of Completeness

Again we have General, Elementary, and Ultimate Problems. These may be stated as follows:

General Completeness Problem: What proportion of the items present in Ideal Check List may be expected to be present also in Sample Check List?

Elementary Completeness Problem: What proportion of the items present in Ideal Check List may be expected to be present also in some designated cell in Sample Check List?

Ultimate Completeness Problem: What proportion of the items in a particular cell in Ideal Check List may be expected to be present also in some designated cell in Sample Check List?

To solve an Ultimate Problem:

From principles already used the proportion to be expected is the same as the value of p alone in an Ultimate Reliability Problem, viz.:

$$\frac{C_{m-n}^{M-N} C_n^N}{C_m^M}.$$

By the use of this formula we may solve the Ultimate Problems for all values of N represented in Ideal Check List and all values of n represented in Sample Check List; some of these solutions will have a value of zero.

For each value of n, if we properly average the solutions of the Ultimate Problems, we obtain a solution of the Elementary Problem for one Sample-Check-List cell in the form of a percentage which may be called the Index of Completeness for the particular Sample-Check-List cell. In securing this average it is necessary to multiply each Ultimate Problem solution by a relative number corresponding to the assumed ratio of number of items in the particular Ideal-Check-List cell to the number of items in all the Ideal-Check-List cells. The source of the assumed relative numbers is the same as that used in evaluating P in the Reliability Problem.

When we have an Index of Completeness for each Sample-Check-List cell we may obtain a Total Index of Completeness for the Sample Check List as a whole by summing the cell-Indices of Completeness of all the cells of the Sample Check List. By an equivalent but preferable method we may divide the last-named result by the sum of the cell-Indices of Completeness of all the cells of the Sample Check List Material (including cell $n = 0$); by this method the C_m^M of the original formula cancels out and so may be disregarded throughout.

A Total Index of Completeness is similarly obtainable for *a* Sample Check List (any Sample Check List selected from *the* Sample Check List) by summing the cell-Indices of Completeness of the appropriate cells. Thus, if desired, a tabulation may be made showing Indices of Completeness for a series of Sample Check Lists differing in extent.

A combined tabulation may show for each of a series of Sample Check Lists its Index of Reliability and its Index of Completeness.

More Complex Case

So far we have considered a validation problem of simple type. In the writer's Check List Materials for Public School Building Specifications[5] a more complex problem was presented, due to the introduction of the concept of the Applicable Case. A Check List for School Building Specifications was developed with a view to its use by school officials or others as an aid in judging proposed school building specifications with reference to their completeness or incompleteness of determination. The position was taken that a new specification ought not to be charged with the omission of a given item unless the building (as repre-

[5] Byrne, L. Check List Materials for Public School Building Specifications. Teachers College, Columbia University. 1931.

sented by the specification) had an Applicable Case for that item. To give a single example, the Check List contains various items relating to the specifying of marble work. It did not seem appropriate to score a specification down for the omission of numerous determinations in marble work, if in fact there was no marble in the building to be determined. This situation is expressed by saying that there are no Applicable Cases for those items.

It seems likely that there are other research problems in which the question ought to be raised whether adequate treatment does not require the introduction of the concept of the Applicable Case. If so a more difficult validation problem is presented than would otherwise be the case.

In the more complex case indicated solution is obtained by making the necessary extensions in the procedures followed for the simple case.

Modifications in Terms and Symbols

M (constant) = total number of specifications in class
D (variable) = number of these specifications containing an Applicable Case for a particular item
N (variable) = number of the latter specifications which also contain the particular item
m (constant) = number of specifications in sample
d (variable) = number of these specifications containing an Applicable Case for the particular item
n (variable) = number of the latter specifications which also contain the particular item

Values of d range from m to 0 by intervals of 1, and those of n range from d to 0 by intervals of 1.

The convention is adopted that values of D range from M downward, and those of N from D downward, by intervals of $\frac{M}{m}$.

(Tabulation) cell will mean the number of items (or the group of items) having a common value of d and a common value of n.

The criterion for membership in the Sample Check List may, for illustrative purposes, be taken as $n \geqq \frac{d}{2}$.

The criterion for membership in the Ideal Check List may, for illustrative purposes, be taken as $N \geqq \frac{D}{2}$.

Problem of Reliability

Following the same principle and line of reasoning as for the simple case we arrive at the same general formula for the solution of an Ultimate Reliability Problem, viz.:

$$\frac{P_R \cdot p_s}{\Sigma Pp}.$$

Meanings of values of P and p are the same as before except that cells must be described respectively in terms of n and d values instead of n values alone, or N and D values instead of N values alone.

P_R is evaluated in the same manner as before, using the new meaning of "cell."

For p_s the evaluation now becomes

$$p = \frac{C_{m-d}^{M-D}\,C_{d-n}^{D-N}\,C_{n}^{N}}{C_{m}^{M}}$$

which through cancellation may be simplified to the working formula

$$p = C_{m-d}^{M-D}\,C_{d-n}^{D-N}\,C_{n}^{N}.$$

The reasoning leading to the denominator C_m^M is unchanged and so this denominator itself remains unchanged. The numerator for the evaluation of p is altered to the extent shown by the consideration that, in producing "favorable" ways, we now have to do with the number of simultaneous possibilities of drawing n specifications from a group of N specifications containing a particular item, drawing $d - n$ specifications from a group of $D - N$ specifications which contain an Applicable Case for this particular item but do not contain this item itself, and of drawing $m - d$ specifications from a group of $M - D$ specifications which contain no Applicable Case for the item.

Problem of Completeness

Following the same principles and line of reasoning as for the simple case we arrive at the following formula for the solution of an Ultimate Completeness Problem:

$$\frac{C_{m-d}^{M-D}\,C_{d-n}^{D-N}\,C_{n}^{N}}{C_{m}^{M}}.$$

By suitable treatment bringing about cancellations the working formula may be reduced to

$$C_{m-d}^{M-D}\,C_{d-n}^{D-N}\,C_{n}^{N}$$

Techniques and Aids in Computation

The present paper is limited to an attempt to explain with adequate fullness the proposed theory of validation for derivative specifications and check lists, and space is lacking in which to exhibit techniques of actual computation. One specimen problem worked out in fairly complete detail, together with remarks on available aids in computation will be found in Appendix A3 in typewritten copies of the writer's "Check List Materials for Public School Building Specifications" on file in the Library of Teachers College, Columbia University; the Appendices are not included in the printed edition.

A NOTE ON SHEPPARD'S CORRECTIONS

By Solomon Kullback

In this note we shall derive a simple relation between the characteristic function of the grouped distribution and the characteristic function of the original continuous distribution, assuming that the frequency curve has high contact with the x-axis at both ends.

If we set $p_s = \int_{x_s - \frac{w}{2}}^{x_s + \frac{w}{2}} f(x)\, dx$, then the characteristic function of the grouped distribution is given by

$$\psi(t) = \sum e^{itx_s} p_s \tag{1}$$

where $i = \sqrt{-1}$. Replacing p_s by its value as given above, we have

$$\begin{aligned} \psi(t) &= \sum e^{itx_s} \int_{x_s - \frac{w}{2}}^{x_s + \frac{w}{2}} f(x)\, dx \\ &= \sum e^{itx_s} \int_{-\frac{w}{2}}^{\frac{w}{2}} f(x + x_s)\, dx \\ &= \int_{-\frac{w}{2}}^{\frac{w}{2}} dx \sum e^{itx_s} f(x + x_s) \\ & \sum e^{itx_s} f(x_s) \int_{-\frac{w}{2}}^{\frac{w}{2}} e^{-itx}\, dx\, . \end{aligned} \tag{2}$$

There is no difficulty about justifying the inversion of the order of integration and summation.

Because of the assumption of high-contact with the axis of x at both ends of the frequency curve, we have

$$\varphi(t) = \int e^{itx} f(x)\, dx = w \sum e^{itx_s} f(x_s) \tag{3}$$

so that

$$\psi(t) = \frac{2}{wt} \sin \frac{tw}{2}\, \varphi(t)\, . \tag{4}$$

This is the desired result, from which there follows the desired moment relations by equating coefficients of $(it)^r$ on both sides of the equation. For example:

$$1 + M_1 it + \frac{M_2}{2!}(it)^2 + \frac{M_3}{3!}(it)^3 + \cdots = \left(1 + \frac{(it)^2 w^2}{4}\frac{1}{3!} + \frac{(it)^4 w^4}{16}\frac{1}{5!} + \cdots\right)$$

$$\left(1 + m_1 it + \frac{m_2}{2!}(it)^2 + \cdots\right)$$

$$= 1 + m_1 it + \frac{(it)^2}{2!}\left(m_2 + \frac{w^2}{12}\right) + \frac{(it)^3}{3!}\left(m_3 + \frac{m_1 w^2}{4}\right) +$$

or

$$M_1 = m_1\,; \qquad M_2 = m_2 + \frac{w^2}{12}\,; \qquad M_3 = m_3 + \frac{m_1 w^2}{4}\,;$$

WASHINGTON, D. C.

THE LIMITING DISTRIBUTIONS OF CERTAIN STATISTICS[1]

By J. L. Doob

There have been many advances in the theory of probability in recent years, especially relating to its mathematical basis. Unfortunately, there appears to be no source readily available to the ordinary American statistician which sketches these results and shows their application to statistics. It is the purpose of this paper to define the basic concepts and state the basic theorems of probability, and then, as an application, to find the limiting distributions for large samples of a large class of statistics. One of these statistics is the tetrad difference, which has been of much concern to psychologists.

I

Let $F(x)$ be a monotone non-decreasing function, continuous on the left, defined at every point of the x-axis, and satisfying the conditions

$$(1) \qquad \lim_{x\to-\infty} F(x) = 0\,, \qquad\qquad \lim_{x\to\infty} F(x) = 1\,.$$

Then the function $F(x)$ is said to be the distribution function of a chance variable $\mathbf{x}$, and $F(x)$ is said to be the probability that $\mathbf{x} < x$. The curve $y = F(x)$ is sometimes called the ogive in statistics. The chance variable $\mathbf{x}$ itself is merely the function x, taken in conjunction with the monotone function $F(x)$.

If $\int_{-\infty}^{\infty} x dF(x)$ exists as an absolutely convergent Stieltjes integral, the value of the integral is called the expectation of $\mathbf{x}$, and will be denoted by $E(\mathbf{x})$.

II

Let $F(x_1, \cdots, x_n)$ be a function defined over n-dimensional space, which is monotone, non-decreasing, continuous on the left in each coordinate if the others are held fast, and which satisfies the conditions

$$(2) \qquad \lim_{x_j\to-\infty} F(x_1, \cdots, x_n) = 0, \quad j = 1, \cdots, n, \qquad \lim_{x_1,\cdots, x_n\to\infty} F(x_1, \cdots, x_n) = 1$$

where in the last limit, $x_1, \cdots, x_n$ become infinite together. Then $F(x_1, \cdots, x_n)$ is said to be the distribution function of a set of chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$, and $F(x_1, \cdots, x_n)$ is said to be the probability that all the inequalities $\mathbf{x}_j < x_j$, $(j = 1, \cdots, n)$, hold simultaneously. It can be shown that the function $F_j(x) = \lim_{\xi_1, \cdots, \xi_{n-1}\to\infty} (\xi_1, \cdots \xi_{j-1}, x, \xi_j, \cdots \xi_{n-1})$ is of the type discussed in §I. The

[1] Research under a grant-in-aid from the Carnegie Corporation.

function $F_j(x)$ is called the distribution function of $\mathbf{x}_j$. The chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are called independent if $F(x_1, \cdots, x_n) = \prod_1^n F_j(x_j)$. The chance variables $\mathbf{x}_1, \cdots, \mathbf{x}_n$ are merely the functions $x_1, \cdots, x_n$ defined over n-dimensional space, taken in conjunction with the function $F(x_1, \cdots, x_n)$.

If $a_1, \cdots, a_n$ are any real numbers, the number $F(a_1, \cdots, a_n)$, the probability that $\mathbf{x}_j < a_j$, $j = 1, \cdots, n$, is also called the probability that a sample $(x_1, \cdots, x_n)$ shall be in the region of n-dimensional space determined by $x_j < a_j$, $j = 1, \cdots, n$. Thus regions of this special type have probabilities attached to them. Using the usual additivity rules, probabilities can be attached to more general regions, and in fact probability can be defined on a collection C of regions including all open sets, closed sets and all sets which can be obtained from them by repeatedly taking sums, products, and complements. (Such point sets are called Borel measurable). The resulting function of point sets is non-negative and completely additive.[2]

If $f(x_1, \cdots, x_n)$ is any function of $x_1, \cdots, x_n$ let E_x be the set of points $(x_1, \cdots, x_n)$ where $f < x$. Suppose that E_x is in the collection C for all values of x, and let $F(x)$ be the probability attached to the set E_x. Then it is readily seen that $F(x)$ has the properties discussed in §I and is therefore the distribution function of a new chance variable $\mathbf{x}$, which will be denoted by $f(\mathbf{x}_1, \cdots, \mathbf{x}_n)$. The chance variable $f(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ is merely the function $f(x_1, \cdots, x_n)$ taken in conjunction with the distribution function $F(x_1, \cdots, x_n)$. (An example is $f(x_1, \cdots, x_n) = x_1 + \cdots + x_n$, determining the chance variable $\mathbf{x}_1 + \cdots + \mathbf{x}_n$.) Suppose that $E(\mathbf{x})$ exists,

$$(3) \qquad E(\mathbf{x}) = \int_{\infty}^{\infty} x\,dF(x).$$

Then it can be shown that the n-dimensional (Lebesgue)-Stieltjes integral

$$(4) \qquad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n)\,dF(x_1, \cdots, x_n)$$

exists and has the value $E(\mathbf{x})$. Conversely the existence of the integral (4) implies that of (3).

If there is a Lebesgue-integrable function $\varphi(x_1, \cdots, x_n)$ such that

$$(5) \qquad F(x_1, \cdots, x_n) = \int_{\infty}^{x_n} \cdots \int_{\infty}^{x_1} \varphi(x_1, \cdots, x_n)\,dx_1 \cdots dx_n,$$

[2] That is, if $p(E)$ is the value of the set function on the set E, and if $E_1, E_2, \cdots$ are point sets with no common points, and which are in C, $p\left(\sum_{m=1}^{\infty} E_m\right) = \sum_{m=1}^{\infty} p(E_m)$.

the function φ is said to be the density function of the distribution. In this case (4) becomes

$$(4') \qquad \int_{-\infty} \cdots \int_{-\infty} f(x_1, \cdots, x_n)\, \varphi(x_1, \cdots, x_n)\, dx_1 \cdots dx_n.$$

The probability attached to a point set E in the collection C is the integral (4) (or (4′) if there is a density function), where $f = 1$ over E and $f = 0$ elsewhere.

III

Let $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \cdots$ be a sequence of chance variables. We suppose that for every integer n, $\mathbf{x}, \mathbf{x}_n$ determine a bivariate distribution. Then it is readily seen from §II that there is a chance variable $|\mathbf{x}_n - \mathbf{x}|$ and therefore that $P\{|\mathbf{x}_n - \mathbf{x}| \leqq \lambda\}$[3] is defined for every number λ. If

$$(6) \qquad \lim_{n \to \infty} P\{|\mathbf{x}_n - \mathbf{x}| \leqq \lambda\} = 1$$

for every positive number λ, the sequence $\mathbf{x}_n$ is said to converge stochastically, or to converge in probability, to $\mathbf{x}$. If α is a constant, $P\{|\mathbf{x}_n - \alpha| \leqq \lambda\}$ is also defined for every number λ, and there is a corresponding definition of stochastic convergence to α. The usual theorems about limits hold: if $\mathbf{x}_n$, $\mathbf{y}_n$ converge stochastically to $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}_n + \mathbf{y}_n$ converges stochastically to $\mathbf{x} + \mathbf{y}$, etc.

An example of stochastic convergence is given by the law of large numbers. Let $\mathbf{x}$ be a chance variable with distribution function $F(x)$ and suppose that $E(\mathbf{x})$, $E(\mathbf{x}^2)$ exist, i.e. that

$$\int_{-\infty}^{\infty} x\,dF(x), \qquad \int_{-\infty}^{\infty} x^2\,dF(x)$$

are absolutely convergent integrals. Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be chance variables whose n-variate distribution function is $\prod_{j=1}^{n} F(x_j)$: we are thus supposing that the variables all have the same distribution and form an independent set. Then $\frac{1}{n}\sum^{n} \mathbf{x}_j$ is a new chance variable, and Tchebycheff's inequality furnishes an immediate proof that $\frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$ converges stochastically to $E(\mathbf{x})$.[4]

[3] Throughout this paper, if γ represents a set of conditions on chance variables, $P\{\gamma\}$ will denote the probability that those conditions are satisfied.

[4] If $\bar{\mathbf{x}}_n = \frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$, $E(\bar{\mathbf{x}}_n) = E(\mathbf{x})$, $E(\bar{\mathbf{x}}_n^2) = \frac{1}{n}E(\mathbf{x}^2)$. Then if λ is any positive number $P\{|\bar{\mathbf{x}}_n - E(\mathbf{x})| > \lambda\} \leqq \frac{E\{[\mathbf{x} - E(\mathbf{x})]^2\}}{n\lambda^2}$ which implies (6).

There is also another kind of convergence, called convergence with probability 1. The sequence $\{\mathbf{x}_n\}$ converges with probability 1 to $\mathbf{x}$ if

$$\lim_{n\to\infty} P\{|\mathbf{x}_n - \mathbf{x}| \leqq \lambda, |\mathbf{x}_{n+1} - \mathbf{x}| \leqq \lambda, \cdots, |\mathbf{x}_{n+p} - \mathbf{x}| \leqq \lambda\} = 1 \tag{7}$$

for every value of $p \geqq 0$, uniformly in $p \geqq 0$ for every positive number λ. If $p = 0$ in (7), (7) becomes (6), so that convergence with probability 1 implies stochastic convergence. Although the converse is not true, if $\{\mathbf{x}_n\}$ is a sequence of chance variables converging stochastically to $\mathbf{x}$, there is a subsequence of $\{\mathbf{x}_n\}$ which converges with probability 1 to $\mathbf{x}$.[5] The usual limit theorems hold here also: if $\mathbf{x}_n$, $\mathbf{y}_n$ converge with probability 1 to $\mathbf{x}$, $\mathbf{y}$, $\mathbf{x}_n + \mathbf{y}_n$ converges with probability 1 to $\mathbf{x} + \mathbf{y}$, etc.

An example of convergence with probability 1 is the following. If in the previous example the hypothesis that $E(\mathbf{x}^2)$ exists is removed, so that only the weaker hypothesis of the existence of $E(\mathbf{x})$ is supposed, the Tchebycheff inequality can no longer be applied, but a different method shows that $\frac{1}{n}\sum_{j=1}^{n} \mathbf{x}_j$ converges with probability 1 (and therefore stochastically) to $E(x)$.[6] This result is known as the strong law of large numbers.

IV

Let $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2, \cdots$ be a sequence of chance variables with distribution functions $F(x), F_1(x), F_2(x), \cdots$ respectively. Then if $\lim_{n\to\infty} F_n(x) = F(x)$ for every value of x, the distribution of $\mathbf{x}_n$ is said to converge to a limiting distribution with distribution function $F(x)$.

As an example, consider the Laplace-Liapounoff theorem. Let $\mathbf{x}_1, \mathbf{x}_2, \cdots$ be a sequence of independent chance variables (i.e. any finite number of them form an independent set) with the same distribution functions, and let $E(\mathbf{x}_n)$, $E(\mathbf{x}_n^2)$ exist. We suppose that $\sigma^2 = E\{[\mathbf{x}_n - E(\mathbf{x}_n)]^2\} > 0$ so that the distribution of $\mathbf{x}_n$ is not merely confined to one point. Then the distribution of

$$n^{-\frac{1}{2}} \sum^{n} [\mathbf{x}_j - E(\mathbf{x}_j)] \tag{8}$$

[5] The theories of probability and of measure are fundamentally identical. Chance variables correspond to measurable functions. Stochastic convergence corresponds to convergence in measure, and convergence with probability 1 corresponds to convergence almost everywhere. The relation between these two types of convergence is discussed (in the terminology of the measure theory) in E. W. Hobson, The Theory of Functions of a Real Variable, second edition Vol. 2, pp. 239–244.

[6] Cf. for instance J. L. Doob, Transactions of the American Mathematical Society, Vol. 36 (1934), pp. 764–765.

converges to a limiting distribution with distribution function[7]

$$(9) \qquad \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2\sigma^2}}\,dx .$$

The convergence of a sequence of n-variate distributions is defined as the convergence of the distribution functions just as above for $n = 1$. Suppose that $(\mathbf{x}_{11}, \cdots, \mathbf{x}_{n1})$, $(\mathbf{x}_{12}, \cdots, \mathbf{x}_{n2})$, $\cdots$ are independent sets of chance variables (i.e. the distribution function of any finite number of sets is the product of the distribution functions of the sets) with the same distribution functions. We suppose that $E(\mathbf{x}_{j1})$, $E(\mathbf{x}_{j1}^2)$ exist, $j = 1, \cdots, n$ and that $\sigma_j^2 = E\{[\mathbf{x}_{j1} - E(\mathbf{x}_{j1})]^2\} > 0$. Then if $\bar{\mathbf{x}}_{jm} = m^{-\frac{1}{2}} \sum_{i=1}^{m} [\mathbf{x}_{ji} - E(\mathbf{x}_{ji})]$, the n-variate distribution of $\bar{\mathbf{x}}_{1m}, \cdots, \bar{\mathbf{x}}_{nm}$ converges to the normal distribution[8] about zero means with variances $\sigma_1^2, \cdots, \sigma_n^2$ and correlation coefficients $\{\rho_{ij}\}$ where $\sigma_i\sigma_j\rho_{ij} = E\{[\mathbf{x}_{i1} - E(\mathbf{x}_{i1})][\mathbf{x}_{j1} - E(\mathbf{x}_{j1})]\}$.

Three lemmas will be needed below in applying these concepts.

LEMMA 1. *If $\{\mathbf{x}_n\}$ is a sequence of chance variables whose distributions approach a limiting distribution and if $\{\mathbf{y}_n\}$ is a sequence of chance variables converging stochastically to* 0, *the sequence $\{\mathbf{x}_n\mathbf{y}_n\}$ converges stochastically to* 0.

For if $F(x)$ is the distribution function of the limiting distribution, and if λ, μ are any positive numbers,

$$\begin{aligned} &P\{|\mathbf{x}_n\mathbf{y}_n| < \lambda\} \geqq P\{|\mathbf{x}_n\mathbf{y}_n| < \lambda, \quad |\mathbf{y}_n| \leqq \mu\} \geqq P\{|\mathbf{x}_n| < \lambda/\mu, \quad |\mathbf{y}_n| \leqq \mu\} \\ (10) \qquad &\geqq P\{|\mathbf{y}_n| \leqq \mu\} - P\{|\mathbf{x}_n| \geqq \lambda/\mu\} = -P\{|\mathbf{y}_n| > \mu\} + P\{|\mathbf{x}_n| < \lambda/\mu\} \\ &\geqq -P\{|y_n| > \mu\} + P\{\mathbf{x}_n < \lambda/\mu\} - P\{\mathbf{x}_n < -\lambda/2\mu\} . \end{aligned}$$

Then, letting n become infinite,

$$(11) \qquad \liminf_{n\to\infty} P\{|\mathbf{x}_n\mathbf{y}_n| < \lambda\} \geqq F(\lambda/\mu) - F(-\lambda/2\mu) .^{9}$$

Letting μ approach 0, $F(\lambda/\mu)$ approaches 1, $F(-\lambda/2\mu)$ approaches 0, and the right hand side becomes 1, as was to be proved.

LEMMA 2. *Let $\{\mathbf{x}_n\}$, $\{\mathbf{y}_n\}$, $\{\mathbf{z}_n\}$ be sequences of chance variables such that the distribution of x_n approaches a limiting distribution with continuous distribution function $F(x)$ and such that the sequences $\{\mathbf{y}_n\}$, $\{\mathbf{z}_n\}$ converge stochastically to* 0, 1 *respectively. Then the distributions of $\{\mathbf{x}_n/\mathbf{z}_n\}$[10] and of $\mathbf{x}_n + \mathbf{y}_n$ approach limiting distributions with the same distribution function $F(x)$.*

[7] A. Khintchine, Ergebnisse der Mathematik, Vol 2, No. 4: Asymptotische Gesetze der Wahrscheinlichkeitsrechnung, pp. 1–8.

[8] Ibid. pp. 11–16.

[9] If $\{a_n\}$ is a sequence of real numbers $\limsup_{n\to\infty} a_n$ is defined as $\lim_{n\to\infty}$ {least upper bound $a_n, a_{n+1}, \cdots$ }, and $\liminf_{n\to\infty} a_n$ is defined as $-\limsup_{n\to\infty}(-a_n)$. A necessary and sufficient condition that the sequence $\{a_n\}$ converge to a limit a is that $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = a$.

[10] Since $\mathbf{z}_n$ converges stochastically to 1, the probability that $\mathbf{z}_n = 0$ approaches 0. The theorem is independent of the way $\mathbf{x}_n/\mathbf{z}_n$ is defined when $\mathbf{z}_n = 0$.

Since $\frac{\mathbf{x}_n}{\mathbf{z}_n} = \mathbf{x}_n + \mathbf{x}_n \frac{1 - \mathbf{z}_n}{\mathbf{z}_n}$ (neglecting the possibility that $\mathbf{z}_n$ may vanish), where the last term converges stochastically to 0 by Lemma 1, it is sufficient to prove the second part of the theorem. If $\epsilon > 0$, and if x is an arbitrary number,

$$(12) \quad P\{\mathbf{x}_n + \mathbf{y}_n < x\} = P\{\mathbf{x}_n + \mathbf{y}_n < x, |\mathbf{y}_n| \leqq \epsilon\} + P\{\mathbf{x}_n + \mathbf{y}_n < x, |\mathbf{y}_n| > \epsilon\}.$$

Since the sequence $\{\mathbf{y}_n\}$ converges stochastically to 0,

$$(13) \qquad \lim_{n \to \infty} P\{\mathbf{x}_n + \mathbf{y}_n < x, |y_n| > \epsilon\} \leqq \lim_{n \to \infty} P\{|\mathbf{y}_n| > \epsilon\} = 0$$

so that in the limit the second term in (12) can be neglected. Moreover

$$(14) \qquad P\{\mathbf{x}_n + \mathbf{y}_n < x, |\mathbf{y}_n| \leqq \epsilon\} \leqq P\{\mathbf{x}_n < x + \epsilon\}.$$

If we let n become infinite and then let ϵ approach 0, (14) becomes

$$(15) \qquad \limsup_{n \to \infty} P\{\mathbf{x}_n + \mathbf{y}_n < x\} \leqq F(x).$$

A similar argument shows that

$$(16) \qquad \liminf_{n \to \infty} P\{\mathbf{x}_n + \mathbf{y}_n < x\} \geqq F(x),$$

and (15), (16) taken together imply that

$$(17) \qquad \lim_{n \to \infty} P\{\mathbf{x}_n + \mathbf{y}_n < x\} = F(x),$$

as was to be proved.

LEMMA 3. *If $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ are chance variables whose distribution has density function*

$$\frac{1}{(2\pi)^2} e^{-\frac{1}{2}(x_1^2 + x_2^2 + x_3^2 + x_4^2)}$$

the distribution of $\mathbf{z} = \mathbf{x}_1\mathbf{x}_2 - \mathbf{x}_3\mathbf{x}_4$ has density function $\frac{1}{2}e^{-|x|}$.

The distribution of $\mathbf{u} = \mathbf{x}_1\mathbf{x}_2$ and that of $\mathbf{v} = -\mathbf{x}_3\mathbf{x}_4$ have the same density function:

$$(18) \qquad \frac{1}{\pi} \int_0^\infty e^{-\frac{x^2}{2t^2} - \frac{t^2}{2}} \frac{dt}{t}.$$

Hence the distribution of $\mathbf{z}$ has density function

$$(19) \qquad \frac{1}{\pi^2} \int_{-\infty}^\infty \int_0^\infty \int_0^\infty e^{-\frac{(x-\lambda)^2}{2t^2} - \frac{t^2}{2} - \frac{\lambda^2}{2\tau^2} - \frac{\tau^2}{2}} d\lambda \frac{dt}{t} \frac{d\tau}{\tau}$$

If we change to polar coördinates: $t = r\cos\theta$, $\tau = r\sin\theta$, and integrate out λ, we obtain

$$\frac{1}{\pi}\int_0^\infty\int_0^{\pi/2} e^{-\frac{x^2}{2r^2}-\frac{r^2}{2}}\,dr\,d\theta = \frac{1}{2}e^{-|x|}.$$

THEOREM 1. *Let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ determine a 4-variate distribution with distribution function $F(x_1, x_2, x_3, x_4)$. Suppose that $E(\mathbf{x}_i)$, $E(\mathbf{x}_i^2)$, $E(\mathbf{x}_i^2\mathbf{x}_j^2)$ exist, $i, j = 1, \cdots, 4$, and suppose that $E(\mathbf{x}_i) = 0$, $E(\mathbf{x}_i^2) = 1$,*[11] *$i, j = 1, 2, 3, 4$. Let $\mathbf{x}_{1j}, \mathbf{x}_{2j}, \mathbf{x}_{3j}, \mathbf{x}_{4j}$ have the same 4-variate distribution as $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$, $j = 1, \cdots, n$, and let the $4n$-variate distribution function of $\{\mathbf{x}_{ij}\}$ be $\prod_{j=1}^{n} F(x_{1j}, x_{2j}, x_{3j}, x_{4j})$. We shall use the following notation (which suppresses the dependence on n):*

$$(20)\qquad \xi_i = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{ik}, \qquad \mathbf{s}_{ij} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{ik}\mathbf{x}_{jk}, \qquad \rho_{ij} = E(\mathbf{x}_i\mathbf{x}_j).$$

Let φ be a function of ξ_i, s_{ij}, defined in a neighborhood N of P: $\xi_i = 0$, $s_{ij} = \rho_{ij}$, which, together with its second partial derivatives is continuous in N. Define $\sigma \geqq 0$ by

$$(21)\qquad \sigma^2 = E\left\{\left[\sum_{i=1}^{4}\frac{\partial\varphi}{\partial\xi_i}\mathbf{x}_i - \sum_{i,j=1}^{4}\frac{\partial\varphi}{\partial s_{ij}}(\rho_{ij} - \mathbf{x}_i\mathbf{x}_j)\right]^2\right\}$$

where the partial derivatives are evaluated at P. Then if $\sigma > 0$, the distribution of $\sqrt{n}\,[\varphi - \varphi(P)]$ (where φ has the arguments ξ_i, $\mathbf{s}_{ij}$) converges to a limiting distribution which is normal with mean 0 and variance σ^2.

To prove this theorem we expand φ in the neighborhood of P, obtaining

$$(22)\qquad \sqrt{n}\,[\varphi - \varphi(P)] = \sum_{i=1}^{4}\frac{\partial\varphi}{\partial\xi_i}\sqrt{n}\,\xi_i - \sum_{i,j=1}^{4}\frac{\partial\varphi}{\partial s_{ij}}\sqrt{n}\,(\rho_{ij} - \mathbf{s}_{ij}) + \mathbf{R}_n$$

where the partial derivatives are evaluated at P, and where $\mathbf{R}_n$ consists of a linear combination of $\sqrt{n}\,\xi_i\xi_j$, $\sqrt{n}\,\xi_i(\rho_{jk} - \mathbf{s}_{jk})$, $\sqrt{n}\,(\rho_{ij} - \mathbf{s}_{ij})(\rho_{kl} - \mathbf{s}_{kl})$, with coefficients which are uniformly bounded as long as ξ_i, $\mathbf{s}_{ij}$ are in the neighborhood N. Now

$$(23)\qquad \lim_{n\to\infty}\xi_i = 0 \qquad\qquad \lim_{n\to\infty}\mathbf{s}_{ij} = \rho_{ij}$$

with probability 1, by the law of large numbers, and as n becomes infinite the distributions of $\sqrt{n}\,\xi_i$, $\sqrt{n}\,(\rho_{ij} - \mathbf{s}_{ij})$ converge to limiting distributions, by the

[11] The hypothesis that $E(\mathbf{x}_i) = 0$ involves no real restriction, since the general case can be reduced to this one by substituting $\mathbf{x}_i - E(\mathbf{x}_i)$ for $\mathbf{x}_i$. The hypothesis that $E(\mathbf{x}_i^2) = 1$ can be met by substituting $\mathbf{x}_i[E(\mathbf{x}_i^2)]^{-\frac{1}{2}}$ whenever $E(\mathbf{x}_i^2) > 0$, which will always be true unless $\mathbf{x}_i = 0$ with probability 1.

Laplace-Liapounoff theorem. Then by Lemma 1, the terms of $\mathbf{R}_n$ converge stochastically to 0. The other terms of $\sqrt{n}[\varphi - \varphi(P)]$ are sums to which the Laplace-Liapounoff theorem can be applied, giving the desired conclusion.

As an example of the application of this theorem, we suppose that φ is a correlation coefficient:

$$\varphi = \frac{\mathbf{s}_{12}}{(\mathbf{s}_{11}\,\mathbf{s}_{22})^{\frac{1}{2}}}, \qquad \varphi(P) = \rho_{12}. \tag{24}$$

Here σ^2 is $E\{[\mathbf{x}_1\mathbf{x}_2 - \frac{1}{2}\rho_{12}(\mathbf{x}_1^2 + \mathbf{x}_2^2)]^2\}$, (which reduces to the familiar result $1 - \rho_{12}^2$ when the bivariate distribution of $\mathbf{x}_1, \mathbf{x}_2$ is normal) and $\sigma = 0$ only when, with probability 1,

$$2\,\mathbf{x}_1\mathbf{x}_2 = \rho_{12}(\mathbf{x}_1^2 + \mathbf{x}_2^2). \tag{25}$$

As a second example we suppose that φ is a tetrad difference:

$$\varphi = \frac{\mathbf{s}_{13}\,\mathbf{s}_{24} - \mathbf{s}_{14}\,\mathbf{s}_{23}}{(\mathbf{s}_{11}\,\mathbf{s}_{22}\,\mathbf{s}_{33}\,\mathbf{s}_{44})^{\frac{1}{2}}}, \qquad \varphi(P) = \rho_{13}\rho_{24} - \rho_{14}\rho_{23}. \tag{26}$$

Here σ^2 becomes

$$\sigma^2 = E\left\{\left[\rho_{24}\mathbf{x}_1\mathbf{x}_3 + \rho_{13}\mathbf{x}_2\mathbf{x}_4 - \rho_{14}\mathbf{x}_2\mathbf{x}_3 - \rho_{23}\mathbf{x}_1\mathbf{x}_4 - \frac{\varphi(P)}{2}\sum_{j=1}^{4}\mathbf{x}_j^2\right]^2\right\} \tag{27}$$

and $\sigma = 0$ only when the quantity in the brackets vanishes with probability 1.

If in either of the two above cases $s_{ij} - \xi_i\xi_j$ is substituted for s_{ij} (i.e. if the deviations from the sample mean, not those from the true mean, are used), the result is unaltered. This is true in general, since $\frac{\partial\varphi}{\partial\xi_i}, \frac{\partial\varphi}{\partial s_{ij}}$ are unaltered at P by this substitution.

There is a well-known δ-method used in statistics to find limiting variances of statistics of the type covered by Theorem 1,[12] and Theorem 1 shows an interpretation which can be given to the results obtained by this method.

We now investigate the necessary modification of Theorem 1 if $\sigma = 0$, i.e. if

$$\sum_{i=1}^{4}\frac{\partial\varphi}{\partial\xi_i}\mathbf{x}_i - \sum_{i,j=1}^{4}\frac{\partial\varphi}{\partial s_{ij}}(\rho_{ij} - \mathbf{x}_i\mathbf{x}_j) = 0 \tag{28}$$

with probability 1. If we assume that φ has continuous third partial derivatives in the neighborhood N, we find that

[12] Examples of the use of this method can be found in T. L. Kelley, Crossroads in The Mind of Man, Stanford University (1928), pp 49 50, and in an article by S. Wright, Annals of Mathematical Statistics, Vol 5 (1934), p. 211.

$$n[\varphi - \varphi(P)] = \frac{n}{2}\sum \frac{\partial^2\varphi}{\partial\xi_i\partial\xi_j}\xi_i\xi_j + \frac{n}{2}\sum_{i,j,k}\frac{\partial^2\varphi}{\partial\xi_i\,\partial s_{jk}}\xi_i(\mathbf{s}_{jk} - \rho_{jk}) \tag{29}$$
$$+ \frac{n}{2}\sum_{i,j,k,l}\frac{\partial^2\varphi}{\partial s_{ij}\,\partial s_{kl}}(\mathbf{s}_{ij} - \rho_{ij})(\mathbf{s}_{kl} - \rho_{kl}) + \mathbf{R}_n'$$

where $\mathbf{R}_n'$ converges stochastically to 0. The second degree terms constitute a quadratic form in $\{\xi_i, \mathbf{s}_{jk} - \rho_{jk}\}$. Now the multivariate distribution of $\{\sqrt{n}\xi_i, \sqrt{n}(\mathbf{s}_{jk} - \rho_{jk})\}$, by the Laplace-Liapounoff theorem, converges to a normal distribution whose variances and correlation coefficients are those of $\mathbf{x}_i$, $\mathbf{x}_j\mathbf{x}_k$. The distribution of $n[\varphi - \varphi(P)]$ thus converges to the distribution of the quadratic form

$$\frac{n}{2}\sum_{i,j}\frac{\partial^2\varphi}{\partial\xi_i\,\partial\xi_j}\alpha_i\alpha_j + \frac{n}{2}\sum_{i,j,k}\frac{\partial^2\varphi}{\partial\xi_i\,\partial s_{jk}}\alpha_i\beta_{jk} + \frac{n}{2}\sum_{i,j,k,l}\frac{\partial^2\varphi}{\partial s_{ij}\,\partial s_{kl}}\beta_{ij}\beta_{kl}, \tag{30}$$

where $\{\alpha_i, \beta_{jk}\}$ have the multivariate distribution just described, unless the quadratic form vanishes identically. This reasoning can be continued, the general result being that there is some power ν of n, if φ is sufficiently regular, such that the distribution of $n^\nu[\varphi - \varphi(P)]$ converges to a limiting distribution.

When $\sigma = 0$ in the second example, unless the distribution of $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ is confined with probability 1 to a 4-dimensional quadric, $\rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = 0$. Equation (29) becomes

$$n[\varphi - \varphi(P)] = \mathbf{s}_{13}\mathbf{s}_{24} - \mathbf{s}_{14}\mathbf{s}_{23} + \mathbf{R}_n'. \tag{29'}$$

Now if $\mathbf{x}_1, \mathbf{x}_2$ are transformed by a linear homogeneous transformation with determinant Δ, it is readily seen that $\mathbf{s}_{13}\mathbf{s}_{24} - \mathbf{s}_{14}\mathbf{s}_{23}$ is multiplied by Δ. The same is true of $\mathbf{x}_3, \mathbf{x}_4$. If $\mathbf{x}_1, \mathbf{x}_2$ are transformed into $\mathbf{x}_1', \mathbf{x}_2'$ so that $E(x_j'^2) = 1$, $E(x_1'x_2') = 0$, the determinant of the transformation is $\pm(1 - \rho_{12}^2)^{-\frac{1}{2}}$. Then transforming each pair $(\mathbf{x}_1, \mathbf{x}_2)$, $(\mathbf{x}_3, \mathbf{x}_4)$ in this way into $(\mathbf{x}_1', \mathbf{x}_2')$, $(\mathbf{x}_3', \mathbf{x}_4')$, the variables $\mathbf{x}_1', \mathbf{x}_2', \mathbf{x}_3', \mathbf{x}_4'$ are uncorrelated. If $\mathbf{s}_{ij}' = \frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_{ik}'\mathbf{x}_{jk}'$,

$$\mathbf{s}_{13}'\mathbf{s}_{24}' - \mathbf{s}_{14}'\mathbf{s}_{23}' = \frac{\mathbf{s}_{13}\mathbf{s}_{24} - \mathbf{s}_{14}\mathbf{s}_{23}}{\pm(1 - \rho_{12}^2)^{\frac{1}{2}}(1 - \rho_{34}^2)^{\frac{1}{2}}}. \tag{31}$$

The limiting distribution of $\mathbf{s}_{13}'\mathbf{s}_{24}' - \mathbf{s}_{14}'\mathbf{s}_{23}'$ is the distribution of $\beta_{13}'\beta_{24}' - \beta_{14}'\beta_{23}'$ where these four chance variables are normally distributed, $E(\beta_{13}') = E(\beta_{24}') = E(\beta_{14}') = E(\beta_{23}') = 0$, $E(\beta_{ij}') = E(\mathbf{x}_i'\mathbf{x}_j')$, $E(\beta_{ij}'\beta_{kl}') = E(\mathbf{x}_i'\mathbf{x}_j'\mathbf{x}_k'\mathbf{x}_l')$. Now if $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ are normally distributed—the most important case for statistical purposes—$\mathbf{x}_1', \mathbf{x}_2', \mathbf{x}_3', \mathbf{x}_4'$ will also be distributed normally, and the vanishing of the correlation coefficients means that the chance variables are independent. If this is true

$$E(\beta_{ij}'^2) = 1, \qquad E(\beta_{ij}\beta_{kl}) = 0, \qquad (\beta_{ij} \neq \beta_{kl}). \tag{32}$$

Evidently, however, $\mathbf{x}_1'$, $\mathbf{x}_2'$, $\mathbf{x}_3'$, $\mathbf{x}_4'$ do not have to be independent to make these equations valid. It is more than sufficient if the pairs $(\mathbf{x}_1, \mathbf{x}_2)$, $(\mathbf{x}_3, \mathbf{x}_4)$ and therefore the pairs $(\mathbf{x}_1', \mathbf{x}_2')$, $(\mathbf{x}_3', \mathbf{x}_4')$ are independent. If (32) is true, the β's are independent, each one being normally distributed with mean 0 and variance 1. Summarizing these results, and using Lemma 3: *if φ is the tetrad difference and if $\rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = 0$, the distribution of $n[\varphi - \varphi(P)]$ converges to a limiting distribution. If in addition the distribution of $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$ is normal, or if the pairs $(\mathbf{x}_1, \mathbf{x}_2)$ $(\mathbf{x}_3, \mathbf{x}_4)$ are independent, this limiting distribution has density function*

$$\frac{c}{2}\, e^{-c|x|}$$

where $c = (1 - \rho_{12}^2)^{-\frac{1}{2}} (1 - \rho_{34}^2)^{-\frac{1}{2}}$.

Wilks has investigated the case where $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$ are normally and independently distributed, and in this case found the exact variance of the tetrad difference as a function of n.[13]

COLUMBIA UNIVERSITY

[13] Proceedings of the National Academy of Sciences, Vol 18, (1932), pp 562-565.

ON THE POSTULATE OF THE ARITHMETIC MEAN

By Richmond T. Zoch

Introduction

Suppose n observations have been made of an unknown quantity. It is desired to know the most probable value of the unknown. When Gauss gave his development of the so-called *Normal Law of Error*, he assumed that *the Arithmetic Mean of the n observations is the most probable value.* The question arises: Can this postulate be justified?

In the excellent book, entitled "Calculus of Observations," by Whittaker and Robinson[1] there is given a proof which purports to deduce the postulate of the Arithmetic Mean from assumptions of a more elementary nature. This proof is not correct.

Since this book has had wide circulation, it is believed that the errors in this proof should be called to the attention of the users of the book. The present paper has been prepared for this purpose. The first part of this paper points out the questionable features of the proof given in Whittaker and Robinson's book. The second part gives some critical comments on the original sources from which Whittaker and Robinson obtained their proof.

Part 1

The assumptions on which Whittaker and Robinson based their proof of the postulate of the Arithmetic Mean are:

Axiom I. The differences between the most probable value and the individual measures do not depend on the position of the null-point from which they are reckoned.

Axiom II. The ratio of the most probable value to any individual measure does not depend on the unit in terms of which the measures are reckoned.

Axiom III. The most probable value is independent of the order in which the measurements are made, and so is a symmetric function of the measures.

Axiom IV. The most probable value, regarded as a function of the individual measures, has one-valued and continuous first derivatives with respect to them.

It is fairly easy to show that if the Arithmetic Mean is the most probable value, then the above four axioms follow as conclusions. The converse, viz. if the above four axioms be assumed then the Arithmetic Mean is the most probable value, however, is not true. That is to say the above assumptions are

[1] The Calculus of Observations by E. T. Whittaker and G. Robinson, Blackie & Son, Ltd., London (1929), pp. 215–217.

necessary conditions, but not sufficient conditions. For, consider the following function of the measures:

$$\frac{\mu_3}{\mu_2} = \frac{\frac{1}{n}\sum_{i=1}^{i=n}(x_i - \bar{x})^3}{\frac{1}{n}\sum_{i=1}^{i=n}(x_i - \bar{x})^2}$$

where $\bar{x}$ is the Arithmetic Mean of the x_i.

Clearly this function is a symmetric function of the measures (x_i) and therefore satisfies Axiom III. If the x_i are each multiplied by k then the Arithmetic Mean $(\bar{x})$ is also multiplied by k and we have

$$\frac{\frac{1}{n}\sum_{i=1}^{i=n}(kx_i - k\bar{x})^3}{\frac{1}{n}\sum_{i=1}^{i=n}(kx_i - k\bar{x})^2} = k\frac{\mu_3}{\mu_2};$$

that is to say, if we multiply the individual measures by k it is the same as multiplying the function $\frac{\mu_3}{\mu_2}$ by k and therefore the ratio of any individual measure to the most probable value (function) does not depend on the unit used. Hence the function $\frac{\mu_3}{\mu_2}$ satisfies Axiom II.

The partial derivative of $\frac{\mu_3}{\mu_2}$ with respect to x_1 is

$$\begin{aligned}\Bigg(\left\{\sum_{i=1}^{i=n}(x_i - \bar{x})^2\right\}&\left[3\left\{\sum_{i=1}^{i=n}(x_i - \bar{x})^2\right\}\left\{-\frac{\partial\bar{x}}{\partial x_1}\right\} + 3(x_1 - \bar{x})^2\frac{\partial x_1}{\partial x_1}\right]\\
&- \left\{\sum_{i=1}^{i=n}(x_i - \bar{x})^3\right\}\left[2\left\{\sum_{i=1}^{i=n}(x_i - \bar{x})\right\}\left\{-\frac{\partial\bar{x}}{\partial x_1}\right\} + 2(x_1 - \bar{x})\frac{\partial x_1}{\partial x_1}\right]\Bigg)\\
&\div \left\{\sum_{i=1}^{i=n}(x_i - \bar{x})^2\right\}^2 = \frac{3\mu_2[(x_1 - \bar{x})^2 - \mu_2] - 2\mu_3(x_1 - \bar{x})}{n\mu_2^2},\end{aligned}$$

since $\frac{\partial\bar{x}}{\partial x_1} = \frac{1}{n}$, and $\sum_{i=1}^{i=n}(x_i - \bar{x}) = 0$. The partial derivatives of $\frac{\mu_3}{\mu_2}$ with respect to each of the x_i are of the same literal form and clearly these partial derivatives are single valued and continuous. Therefore the function $\frac{\mu_3}{\mu_2}$ satisfies Axiom IV.

Now it can be shown that if h be added to each x_i, then the function $\frac{\mu_3}{\mu_2}$ is unchanged and hence this function does not satisfy Axiom I. (It should be noted that the function $\frac{\mu_3}{\mu_2}$ is invariant under the transformation specified by

Axiom I.) However, consider the function $\bar{x} + a\frac{\mu_3}{\mu_2} \equiv f$, where a is a constant independent of the x_i. Clearly, f satisfies all of the four axioms.

Thus a function, distinct from the Arithmetic Mean, has here been exhibited which satisfies the four axioms given in Whittaker and Robinson's book. Hence, these four axioms are not sufficient to establish the postulate of the Arithmetic Mean. The question arises: Where is the proof given by Whittaker and Robinson lacking in rigor? The proof given is essentially as follows. (No part of the proof given by Whittaker and Robinson is here omitted; in fact, for the sake of rigor and careful reasoning, further explanations are given and the various steps are numbered.)

(1) Suppose the most probable value is expressed in terms of the n measures $x_1, x_2, \cdots, x_n$ by the function $\phi(x_1, x_2, \cdots, x_n)$; that is to say the most probable value is some function, ϕ, of the observations, or: the most probable value $\equiv \phi(x_1, x_2, \cdots, x_n)$.

(2) By the theorem of the mean value in the differential calculus, which by Axiom IV is applicable, we have $\phi(kx_1, kx_2, \cdots, kx_n) =$

$$\phi(0, 0, \cdots, 0) + kx_1\left[\frac{\partial\phi}{\partial x_1}\right] + \cdots + kx_n\left[\frac{\partial\phi}{\partial x_n}\right],$$

where the square brackets denote that every x_i is to be replaced by θkx_i where θ lies between 0 and 1.

(3) By Axiom II, the left hand side $= k\phi(x_1, x_2, \cdots, x_n)$.

(4) By the continuity of ϕ, postulated in Axiom IV the equation $\phi(kx_1, kx_2, \cdots, kx_n) = k\phi(x_1, x_2, \cdots, x_n)$ must hold in the limit when k is 0, that is $\phi(0, 0, \cdots, 0) = 0$.

(5) We now have

$$k\phi(x_1, x_2, \cdots, x_n) = kx_1\left[\frac{\partial\phi}{\partial x_1}\right] + \cdots + kx_n\left[\frac{\partial\phi}{\partial x_n}\right],$$

or on dividing by k,

$$\phi(x_1, x_2, \cdots, x_n) = x_1\left[\frac{\partial\phi}{\partial x_1}\right] + \cdots + x_n\left[\frac{\partial\phi}{\partial x_n}\right].$$

(6) In this last equation let $k \to 0$: then each of the quantities $\left[\frac{\partial\phi}{\partial x_i}\right]$ tends to a value which is independent of the x's and we can write $\phi(x_1, x_2, \cdots, x_n) = c_1x_1 + \cdots + c_nx_n$ where the c's are independent of the x's.

(7) By Axiom III the c's must all be equal, so

$$\phi(x_1, x_2, \cdots, x_n) = c(x_1 + x_2 + \cdots + x_n).$$

(8) From Axiom I we have

$$\phi(x_1 + h, x_2 + h, \cdots, x_n + h) = \phi(x_1, x_2, \cdots, x_n) + h.$$

(9) If in this last equation we let the x_i all approach zero then we have $cnh = h$ and therefore $c = \frac{1}{n}$ and finally

$$\phi(x_1, x_2, \cdots, x_n) = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

which states that $\phi \equiv$ the most probable value $=$ the Arithmetic Mean.

It should be noted that the first six steps involve only Axioms II and IV. Of these first six steps the second and sixth are questionable.

The sixth step involves the tacit assumption that the partial derivatives are functions of k. These partial derivatives are not necessarily functions of k and the example given above, viz, $f \equiv \bar{x} + a\frac{\mu_3}{\mu_2}$ is a function whose partial derivatives are independent of k; in fact no function of the form

$$F \equiv \bar{x} + \sum_{j=3}^{j=\infty} a_j \frac{\sum_{i=1}^{i=n} (x_i - \bar{x})^j}{\sum_{i=1}^{i=n} (x_i - \bar{x})^{j-1}}$$

will satisfy the tacit assumption involved in the sixth step; nor is F the most general function which will not satisfy the tacit assumption, thus take for example

$$\bar{F} \equiv \bar{x} + \frac{a\mu_3\mu_4}{b\mu_2\mu_4 + c\mu_3^2}.$$

Consider now the second step. Take the function $\phi(y_1, y_2, \cdots, y_n) = k\phi(x_1, x_2, \cdots, x_n)$. Then, by Axiom II, we have $y_i = kx_i$. Apply the Theorem of the Mean Value to $\phi(y_i)$ instead of $\phi(x_i)$. Then $\phi(y_1, y_2, \cdots, y_n) = \phi(0, 0, \cdots, 0) + y_1\left[\frac{\partial\phi}{\partial y_1}\right] + \cdots + y_n\left[\frac{\partial\phi}{\partial y_n}\right]$. Now if we replace y_i by kx_i we obtain the equation given in the second step except that the square brackets are now of the form $\left[\frac{\partial\phi(kx_1, kx_2, \cdots, kx_n)}{\partial(kx_i)}\right]$ and not $\left[\frac{\partial\phi}{\partial x_i}\right]$ as given by Whittaker and Robinson. It is difficult to decide whether by $\left[\frac{\partial\phi}{\partial x_i}\right]$ Whittaker and Robinson mean

$$\left[\frac{\partial\phi(kx_1, kx_2, \cdots, kx_n)}{\partial x_i}\right] \text{ or } \left[\frac{\partial\phi(x_1, x_2, \cdots, x_n)}{\partial x_i}\right].$$

These last two expressions are not equal. To make the second step more clear it is necessary to demonstrate that

$$\left[\frac{\partial\phi(kx_1, kx_2, \cdots, kx_n)}{\partial(kx_i)}\right] = \left[\frac{\partial\phi(x_1, x_2, \cdots, x_n)}{\partial x_i}\right],$$

and this has not been done. In order to demonstrate this equality further use must be made of Axiom II. It appears that the questionable features of the second step may be overcome by starting with the equation implied by Axiom II, thus

$$\phi(kx_1, kx_2, \cdots, kx_n) = k\phi(x_1, x_2, \cdots, x_n);$$

in other words ϕ is a homogeneous function of degree 1. Therefore use can be made of Euler's Theorem on homogeneous forms. In this way we obtain:

$$\phi = \sum_{i=1}^{i=n} x_i \frac{\partial \phi}{\partial x_i}$$

which is an abbreviation of the last equation given in the fifth step.

Now, making further use of Axiom II we have:

$$\frac{\partial \phi(kx_1, kx_2, \cdots, kx_n)}{\partial (kx_i)} = \frac{\partial}{\partial (kx_i)} k\phi(x_1, x_2, \cdots, x_n) = k \cdot \frac{1}{k} \cdot \frac{\partial}{\partial x_i} \phi(x_1, x_2, \cdots, x_n) .$$

It follows that

$$\frac{\partial \phi(x_1, x_2, \cdots, x_n)}{\partial x_i} = \frac{\partial \phi(kx_1, kx_2, \cdots, kx_n)}{\partial (kx_i)} .$$

From this development we conclude that for any function whatever which satisfies Axiom II the last equation of the fifth step cannot possibly involve k.

In order to overcome the defect in the sixth step it is necessary to make a more restrictive assumption. If in place of Axiom IV, we assume that "*The most probable value, regarded as a function of the individual measures, has first partial derivatives with respect to them which are constant,*" then the equation given in the sixth step can be rigorously established.

After the equation of the sixth step is rigorously established there remains an objection in the seventh step. The axioms do not explicitly state that the n observations must be functionally independent. Therefore suppose the x_i are functionally dependent according to the relation $x_i = y_i z$ where the y_i are all constant. Then the function $f \equiv \bar{x} + \frac{\mu_3}{\mu_2}$ will have partial derivatives with respect to the x_i which are unequal and constant; yet at the same time the function f is a symmetrical expression of the n variables.

Hence in order to establish the postulate of the Arithmetic Mean along the lines followed by Whittaker and Robinson it is necessary to make another restrictive assumption slightly different from that proposed in the last paragraph but one, and assume (in addition to Axioms I and II) that *the function has partial derivatives with respect to the x_i which are equal.*

Part 2

The first original paper consulted was one by Schiaparelli.[2] In this paper nine propositions are presented four of which are also called lemmas. From a strict mathematical point of view the four propositions which Schiaparelli calls lemmas are really postulates. Schiaparelli discusses these four lemmas at length; three of these lemmas are the first three axioms given in Whittaker and Robinson's book. The fourth one is: "When, in the function ϕ, all the variables (x_i) take the same value a, the function itself becomes equal to a," (This, as a matter of fact, is the definition of an average).

In his discussion of these lemmas, which are based partly on practical and partly on philosophical grounds, Schiaparelli points out that they are justified from the practical or statistical nature of the problem involved in arriving at the most probable value (Schiaparelli uses the term "true value") of a set of observations. In the present writer's opinion, these discussions are the most excellent part of Schiaparelli's paper. These discussions are even more significant in view of the fact that the later writers on this subject make no attempt whatsoever to justify the use of their postulates.

Schiaparelli remarks that we should have no reason for not expecting that a small change in a single observation should produce a small change in the function ϕ; but he does not make this remark in the form of an explicit postulate. This could have been done and, moreover, such a postulate of continuity could be justified from the practical nature of the problem. It seems that a more elegant procedure would have been to deduce the continuity of the function and its derivatives from Axioms I and II. It will be shown later that this is possible. From his remark on the continuity of the function, Schiaparelli concludes that the partial derivatives of ϕ with respect to the x_i exist and are continuous. His method of arriving at this conclusion is not valid, for it is well known that an arbitrarily assumed function may be everywhere continuous and yet possess a derivative at no point.

Schiaparelli's Proposition III states: "When in the function ϕ all the x_i take the same value, then the $\dfrac{\partial \phi}{\partial x_i}$ become equal to each other." This Proposition is false. To show this, consider the function

$$f \equiv \bar{x} + \frac{\mu_3}{\mu_2},$$

where the

$$\frac{\partial f}{\partial x_i} = \frac{1}{n} + \frac{3\mu_2[(x_i - \bar{x})^2 - \mu_2] - 2\mu_3(x_1 - \bar{x})}{n\mu_2^2}$$

[2] Giovanni Schiaparelli—Come si possa giustificare l'uso della media aritmetica nel calcolo die risultati d'osservazione, Rendiconti Reale Instituto Lombardo di Scienze e lettere, Vol. XL (1907), pp 752–764.

Now, when the x_i all approach a then both f and $\frac{\partial f}{\partial x_i}$ become indeterminate forms. However, in this case f takes an indeterminate form which can be evaluated and it can be shown that $\frac{\mu_3}{\mu_2}$ will always have the value zero, i.e., f will have the value a when all the $x_i = a$; while the $\frac{\partial f}{\partial x_i}$ can take any value whatever and in general the $\frac{\partial f}{\partial x_i}$ will not be equal when the $x_i \to a$. To illustrate: Consider the observations $y_1 = 1$, $y_2 = 3$, $y_3 = 4$ then $\bar{y} = 8/3$ and $\mu_2 = 14/9$ and $\mu_3 = -20/27$ whence $f = 8/3–10/21$. Now assume that these three observations all approach 2 in a certain way, i.e., let $x_i = 2 + (y_i - 2)z$. Then $\bar{x} = 2 + (\bar{y} - 2)z = 2 + (2/3)z$.

$$\mu_2(x_i) = z^2 \frac{1}{n} \Sigma (y_i - \bar{y})^2 = (14/9)z^2$$

and

$$\mu_3(x_i) = z^3 \frac{1}{n} \Sigma (y_i - \bar{y})^3 = (-20/27)z^3$$

whence $f = 2 + (2/3)z - (10/21)z$. Clearly as $z \to 0$ the $x_i \to 2$ and $f \to 2$. However,

$$\left.\frac{\partial f}{\partial x_1}\right]_{x_1=2+(y_1-2)z} = \frac{1}{3} + \frac{131}{294},$$

$$\left.\frac{\partial f}{\partial x_2}\right]_{x_2=2+(y_2-2)z} = \frac{1}{3} - \frac{253}{294},$$

$$\left.\frac{\partial f}{\partial x_3}\right]_{x_3=2+(y_3-2)z} = \frac{1}{3} + \frac{122}{294}.$$

Thus the $\frac{\partial f}{\partial x_i}$ are not functions of z and as the $x_i \to 2$ the $\frac{\partial f}{\partial x_i}$ remain constant and unequal.

From his conclusion that the derivatives of ϕ exist and from Axiom I, Schiaparelli obtains the equation, $\sum_{i=1}^{i=n} \frac{\partial \phi}{\partial x_i} = 1$, (this equation being his Proposition V) in the following way: Since the derivatives of ϕ exist, then by the Theorem of the mean value,

$$\phi(x_1 + h, x_2 + h, x_3 + h, \cdots, x_n + h)$$
$$= \phi(x_1, x_2, \cdots, x_n) + h\left(\frac{\partial \phi}{\partial x_1} + \frac{\partial \phi}{\partial x_2} + \cdots + \frac{\partial \phi}{\partial x_n}\right). \quad \text{(A)}$$

By Axiom I:

$$\phi(x_1 + h, x_2 + h, \cdots, x_n + h) = \phi(x_1, x_2, \cdots, x_n) + h.$$

Whence $\sum_{i=1}^{i=n} \frac{\partial \phi}{\partial x_i} = 1$. Now this equation is correct but the above proof of it is not convincing. Clearly, according to the Theorem of the mean value, in equation (A) it is necessary to replace each x_i in the $\frac{\partial \phi}{\partial x_i}$ by θx_i where θ is between 0 and 1.

Schiaparelli's Proposition VII states in effect that the $\frac{\partial \phi}{\partial x_i}$ are invariant under the transformation $x_i' = x_i + h$ where h is constant, and his Proposition IX states that the $\frac{\partial \phi}{\partial x_i}$ are invariant under the transformation $x_i' = kx_i$ where k is a constant. These two propositions are correct and are correctly established. Making use of his Propositions III (which is false), V, VII and IX, Schiaparelli proceeds to the establishment of the postulate of the Arithmetic Mean, as follows:

Let $a = \phi(x_i)$. As the x_i vary, then a varies but for a particular set of x_i then a is a constant. Now by Axiom I we have

$$a + (m-1)\,a = \phi(x_1 + (m-1)a,\, x_2 + (m-1)a,\, \cdots,\, x_n + (m-1)a) = m\,a$$

for all values of $m > 1$. Then by Axiom II:

$$a = \phi\left(\frac{x_1 + (m-1)\,a}{m}, \frac{x_2 + (m-1)\,a}{m}, \cdots, \frac{x_n + (m-1)\,a}{m}\right)$$
$$= \phi\left(\frac{x_1 - a}{m} + a, \frac{x_2 - a}{m} + a, \cdots, \frac{x_n - a}{m} + a\right).$$

And by Propositions VII and IX, the $\frac{\partial \phi}{\partial x_i}$ are unchanged during the above transformations. Hence the last equation is true when $m \to \infty$ and by Proposition III (false) the $\frac{\partial \phi}{\partial x_i} = \frac{1}{n}$ as when $m \to \infty$, $\phi(x_i) = a$. In this final proof Schiaparelli gives a geometric illustration of each step.

It is both interesting and strange to know that in closing his paper Schiaparelli does not claim that the Arithmetic Mean is the only function which will satisfy all of his postulates. In fact he himself points out that the function ϕ, implicitly defined by the equation $\sum_{i=1}^{i=n} (\phi - x_i)^m = 0$ where m is an odd integer > 1 will satisfy all of his postulates. Furthermore he points out that this function will not satisfy his Proposition III. Schiaparelli's object was to establish the postulate of the Arithmetic Mean without any appeal to the concept of probability. To accomplish this he made four assumptions each of which he justified by *a priori* reasoning. Then he proceeded with the above proof. Why he should have been satisfied with his own proof after perceiving the function defined by $\sum^{i=n} (\phi - x_i)^m = 0$ is hard to understand.

The second paper[3] consulted was also by Schiaparelli. It is merely an abridged form of the one just discussed. Schiaparelli wrote two earlier papers on this same subject (altogether Schiaparelli wrote four papers on it) but it was inferred from the footnotes in his paper, which has just been discussed at length, that it contained all of the material of the two earlier papers with which he himself was satisfied. Therefore Schiaparelli's two earlier papers were not consulted.

The third paper consulted was that by Broggi.[4] Broggi states that the purpose of his paper is to establish the postulate of the Arithmetic Mean by purely analytic methods which are more brief than Schiaparelli's method. Broggi words the assumptions upon which he bases his proof as follows:

1. ϕ is a symmetric function of its n variables;
2. The partial derivatives are single-valued and finite;
3. We have $\phi(kx_1, kx_2, \cdots, kx_n) = k\phi(x_1, x_2, \cdots, x_n)$;
4. We have $\phi(x_1 + h, x_2 + h, \cdots, x_n + h) = \phi(x_1, x_2, \cdots, x_n) + h$, that is to say for 2:

$$\frac{\partial\phi}{\partial x_1} + \frac{\partial\phi}{\partial x_2} + \cdots + \frac{\partial\phi}{\partial x_n} = 1\,. \tag{a}$$

Broggi does not explain why he used the postulate 2 but presumably it was in order to exclude the function defined by $\sum_{i=1}^{i=n} (\phi - x_i)^m = 0$. Consider the special case where $m = 3$. Then $n\phi^3 - 3\phi^2\,\Sigma x_i + 3\phi\,\Sigma x_i^2 - \Sigma x_i^3 = 0$. Let $p = 3\left(\frac{1}{n}\,\Sigma x_i^2 - \bar{x}^2\right)$ and $q = \frac{3\bar{x}}{n}\,\Sigma x_i^2 - 2\bar{x}^3 - \frac{1}{n}\,\Sigma x_i^3$. Also put $R = (p/3)^3 + (q/2)^2$ and let A be the *real* cube root of $-q/2 + \sqrt{R}$ and B be the *real* cube root of $-q/2 - \sqrt{R}$. Then the three branches of ϕ can be explicitly written

$$\phi_1 = A + B + \bar{x}$$

$$\phi_2 = \omega A + \omega^2 B + \bar{x}$$

$$\phi_3 = \omega^2 A + \omega B + \bar{x}$$

where ω and ω^2 are the two complex cube roots of unity. Now while ϕ does not satisfy the postulate that the function be single valued, ϕ_1 satisfies this postulate as well as all the others and so does ϕ_2 and also ϕ_3. Hence, Broggi's failure to comment at length on the function $\sum_{i=1}^{i=n} (\phi - x_i)^m = 0$ is unsatisfying. As a matter of fact Broggi fails to point out any of the defects of Schiaparelli's

[3] Giovanni Schiaparelli—Come si possa giustificare l'uso della media aritmetica nel calcolo delle misure, senza fare alcuna ipotesi sulla legge di probabilità degli errori accidentali, Astronomische Nachrichten, Band 176 (1907) pp. 206-212.

[4] Ugo Broggi—Sur Le Principe De La Moyenne Arithmetique, L'Enseignement Mathematique, XI (1909) pp. 14–17.

paper, with the possible exception that he shows Schiaparelli's postulate which states $\phi = a$ when each of the $x_i = a$ to be a consequence of Axioms I and II. This is done so casually that it makes one wonder whether Broggi really was aware of the fact that Schiaparelli's postulates are not independent.

Broggi proves the Lemma: "A homogeneous function of the first degree which is a solution of the equation of partial derivatives (a) is an integral function." This Lemma is correct and is correctly proved but its wording is apt to be misleading; in fact it appears that its true meaning was not clear to Broggi himself. For, while the function ϕ cannot be of the form $\frac{\psi}{\chi}$ where ψ is a homogeneous function of the p^{th} degree which satisfies Axiom I and χ a homogeneous function of the $(p - 1)^{\text{th}}$ degree which also satisfies Axiom I, the Lemma does not mean and Broggi has not proved that ϕ cannot be of the form $\phi = \Omega + \frac{\psi}{\chi}$ where Ω is an integral function satisfying Axioms I and II and ψ and χ are homogeneous functions of the p^{th} and $(p - 1)^{\text{th}}$ degrees respectively which are *invariant* under the transformation specified in Axiom I. By reason of this oversight, Broggi concludes that any function satisfying Axioms I and II must be linear in its n variables, a conclusion which is erroneous.

The fourth paper consulted was that by Schimmack.[5] Schimmack's paper is in three sections. The first section contains the proof which is essentially that which Whittaker and Robinson give. In the second section Schimmack gives a different proof, from a set of new postulates. The new set of postulates is:

Axiom I′ = Axiom I.

Axiom II′—The most probable value is independent of the sense of direction of the scale upon which the observed values (and the most probable value) are reckoned, that is to say,

$$\phi(-x_1, -x_2, \cdots, -x_n) = -\phi(x_1, x_2, \cdots, x_n).$$

Axiom III′ = Axiom III.

Axiom IV′—If from n observed values, the most probable value be computed and if one obtains an additional observed value then the most probable value of the $n + 1$ observed values is the same as the most probable value of $n + 1$ quantities consisting of the initial most probable value counted n times and the $(n + 1)^{\text{th}}$ observed value, namely:

$$\phi_{n+1}(x_1, \cdots, x_{n+1}) = \phi_{n+1}(\phi_n, \cdots, \phi_n, x_{n+1}).$$

In explaining the object of this second section, Schimmack says that postulating the existence of the derivatives (Axiom IV) seems unjustified and ought to be avoided and only such axioms made which the intrinsic character of the problem justifies. In connection with this statement of Schimmack's it appears that the intrinsic character of the problem certainly does not justify Axiom IV′. In

[5] Rudolf Schimmack—Der Satz vom arithmetischen Mittel in axiomatischer Begründung, Mathematische Annalen, Band 68 (1909) pp. 125–132, 304.

fact, Axiom IV′ appears to be quite artificial. Moreover, Schimmack does not attempt to justify Axiom IV′ by *a priori* reasoning as Schiaparelli does for Axioms I, II, and III. While, if the Arithmetic Mean is the most probable value, Axiom IV′ follows, since it is a property of the Arithmetic Mean, it does not seem to be in keeping with the intrinsic character of the problem to use this property as a starting point for later deductions.

As regards Schimmack's objections to Axiom IV, all of the conditions specified by it can be deduced from the first two Axioms except that the derivatives must be single-valued. To show that this is true, consider an arbitrary function which satisfies Axioms I and II. Let this function be $\phi(x_1, x_2, \cdots, x_n)$. We do not know that ϕ is continuous or that ϕ has any derivatives. All we assume is that ϕ satisfies the first three Axioms and it is here proven that ϕ must be continuous and have continuous partial derivatives. By Axiom I we can give increments to the x_i; hence we give each x_i the same increment, Δx, and then subtract ϕ and we have: $\phi(x_1 + \Delta x, x_2 + \Delta x, \cdots, x_n + \Delta x) - \phi(x_1, x_2, \cdots, x_n) = \Delta\phi$ but by Axiom I, $\Delta\phi = \Delta x$. Therefore $\frac{\Delta\phi}{\Delta x} = 1 = \frac{d\phi}{dx}$. In other words, the total derivative of ϕ exists and is constant. Therefore the total derivative of ϕ is continuous. But since the total derivative exists, all of the partial derivatives exist. By Axiom II, ϕ is a homogeneous function of the first degree. Applying Euler's Theorem for homogeneous forms, we have $\phi = x_1 \frac{\partial\phi}{\partial x_1} + x_2 \frac{\partial\phi}{\partial x_2} + \cdots + x_n \frac{\partial\phi}{\partial x_n}$. Since the total derivative of ϕ is everywhere continuous, ϕ is also everywhere continuous. Thus, the right hand side of the above equation is everywhere continuous and each partial derivative is therefore everywhere continuous.

As regards that part of Axiom IV which requires the $\frac{\partial\phi}{\partial x_1}$ to be single valued, it would seem more satisfactory to postulate that the function ϕ is single-valued, for the single-valuedness of a derivative does not insure the single-valuedness of the integral while the single-valuedness of a function does insure the single-valuedness of the derivative where the derivative exists.

In the third section of his paper, Schimmack shows Axioms I, II, III, and IV to be independent, and likewise Axioms I, II′, III and IV′.

Schimmack does not mention any of the questionable features of Schiaparelli's and Broggi's papers.

The fifth paper consulted was that by Suto.[6] Suto's assumptions are:

1°. $\phi(x, x, \cdots, x) = x$ (This is Schiaparelli's).

2°. $\phi(x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n) - \phi(x_1, x_2, \cdots, x_n)$ depends on the values of $y_1, y_2, \cdots, y_n$ only.

3°. = Axiom III = Axiom III′.

[6] Onosaburo Suto—Law of the Arithmetical Mean, Tohoku Mathematical Journal, Vol. 6 (1914) pp. 79–81.

Suto says he believes these assumptions to be more simple and natural than Schimmack's Axioms I′–IV′. However, assumption 2° appears to be quite artificial and very restrictive. Suto does not even attempt to justify it by *a priori* reasoning.

Suto shows his three Axioms to be independent. It is interesting to know that Suto has established the postulate of the Arithmetic Mean rigorously using only three postulates while Schiaparelli, Broggi and Schimmack failed using four postulates. In this connection it should be observed that when Axiom IV as given by Whittaker and Robinson is replaced by "The most probable value, regarded as a function of the individual measures, has first partial derivatives with respect to them which are equal" as suggested at the end of Part 1, then Axiom III can be deduced as a consequence of Axioms I, II and the reworded Axiom IV, so that three Axioms only are sufficient to deduce the postulate of the Arithmetic Mean. However, it would be difficult to justify the reworded Axiom IV from the nature of this problem of the Arithmetic Mean.

Suto does not point out any of the defects of the preceding papers.

The last paper consulted was that by Beetle.[7] It deals with the third section of Schimmack's paper. Beetle also fails to point out any of the defects of the preceding papers.

Conclusion

The postulate of the Arithmetic Mean can be rigorously established, without the use of the concept of probability, if sufficiently restrictive assumptions are made. The writers making sufficiently restrictive assumptions have failed to justify the use of them. Several proofs of the postulate of the Arithmetic Mean are clearly erroneous. The existing attempts to establish the postulate of the Arithmetic Mean without any appeal to the concept of probability are, therefore, unsatisfactory.

Acknowledgment

This paper was prepared under the direction of Dr. Frank M. Weida of the George Washington University.

THE GEORGE WASHINGTON UNIVERSITY,
WASHINGTON, D. C.

[7] R. D. Beetle—On the complete independence of Schimmack's postulates for the Arithmetic Mean, Mathematische Annalen, Band 76 (1915) pp. 444–446.

THE SHRINKAGE OF THE BROWN-SPEARMAN PROPHECY FORMULA

By Robert J. Wherry

At the recent meeting of the Conference on Individual Psychological Differences held in Washington, Dr. Clark Hull of Yale University called attention to the fact that the much used Brown-Spearman formula involves, or leads to, if used without regard to certain limitations, a certain over optimism.[1] In other words, if only this formula is taken into account, one would assume that the mere increasing in length of a test would automatically and, with continued increases in length, *indefinitely* continue to increase its reliability or validity.

On the other hand, we know that the greater the number of test units the greater the shrinkage between the predicted and actually obtained value. At least we know this to be true when the value in question is a multiple correlation coefficient and the test units are independent variables. Hull raised the question as to whether or not the same fact might be true of the figures predicted by the Brown-Spearman formula. It is the purpose of this article to show that this shrinkage does occur, and that the Wherry-Smith shrinkage formula[2] satisfactorily predicts this shrinkage.

A quick review of the nature of the two formulae (the Brown-Spearman and the Wherry-Smith formulae) will at once show the importance of the discussion. The Brown-Spearman formula, as applied to the predicting of reliability, reads as follows,

$$R = \frac{M\, r_{11}}{1 + (M - 1)\, r_{11}}, \tag{1}$$

where R = the predicted reliability,
r_{11} = the discovered reliability,
and M = the number of times the test is lengthened. Thus the test provides that the predicted reliability (R) will increase with each increase in M, but it is to be noted that the increase in R decreases with each increase in M as the value of R approaches its limit of plus one.

On the other hand the Wherry-Smith formula, which reads,

$$\bar{R}^2 = \frac{(N - 1)R^2 - (M - 1)}{N - M}, \tag{2}$$

where $\bar{R}$ = the predicted value of the correlation,
R = the discovered correlation,
M = the number of independent variables

and N = the statistical population (the number of cases), provides that, for each increase in M, the shrinkage in $\bar{R}$ as compared with R increases. Thus, if

TABLE I

Correlations Observed and Theoretical (Based upon Observed Means)

(N = 37 throughout)

M	Observed average	Correlation predicted		Error	
		Br.-Sp.	Wherry	Br.-Sp.	Wherry
	(Trait 1)				
1	.290				
5	.728	.671	.618	−.057	−.110
10	.717	.803	.726	.086	.009
15	.754	.860	.758	.106	.004
20	.805	.891	.825	.087	.020
30	.936	.925	.509	−.011	−.427
	(Trait 5)				
1	.419				
5	.736	.783	.751	.047	.015
10	.845	.878	.834	.033	−.011
15	.887	.915	.856	.028	−.031
20	.877	.935	.856	.058	−.021
30	.876	.956	.745	.080	−.131
	(Trait 10)				
1	.354				
5	.479	.733	.692	.254	.213
10	.717	.846	.788	.129	.071
15	.852	.892	.816	.040	−.036
20	.636	.915	.822	.279	.186
30	.805	.943	.655	.138	−.150
	(All Traits)				
1	.320				
20	.898	.904	822	.006	−.076
30	.872	.933	.576	.061	−.296

we assume that the M's in the two formulae are analogous, i.e., if we assume the Wherry-Smith formula to be applicable to the Brown-Spearman formula, we see that as M increases the Brown-Spearman formula adds a decreasing incre-

ment while the Wherry-Smith formula provides that an increasing decrement be subtracted, thus eventually we arrive at a point where by further increasing the length of the test we will decrease rather than increase the size of the reliability coefficient.

If our hypothesis be true, we must, then, in order to predict the correct value of $\bar{R}$, substitute the value of equation (1) in equation (2). Doing this we have

$$\bar{R}^2 = \frac{(N-1)M^2r_{11}^2 - (M-1)^3r_{11}^2 - 2(M-1)^2r_{11} - (M-1)}{(N-M)[1 + 2(M-1)r_{11} + (M-1)^2r_{11}^2]} \tag{3}$$

which would then be the form in which the Brown-Spearman formula should be used in predicting reliability corrected for chance error by the Wherry-Smith

TABLE II

Error in Predicting Reliability (Based upon Observed Means)

Error	Brown-Spearman	Wherry
over .210	2	1
.151– .210		1
.091– .150	3	
.031– .090	8	1
–.029– .030	3	6
–.089– –.030	1	3
–.149– –.090		3
–.209– –.150		
below –.209		2

TABLE III

Rietz Criteria of Normality Applied to Results from Means

Criterion	Normal	Brown-Spearman	Wherry
u_1	0	.074	–.032
β_1	0	.561	–.283
β_2	3	2.008	3.180

formula. The same result can of course be secured by applying the formulae consecutively.

In order to test the formula (3), the writer has applied it to some empirical data. A recent article by H. H. Remmers of Purdue University furnishes the needed data. Remmers study dealt with the increase in reliability due to increase in the number of judgments of certain traits of college professors.[3] His results, together with the results of applying formula (3) to the data are shown in Table I.

An inspection of Table I shows at once that while the Brown-Spearman

formula gives results which are consistently too large (15 out of 17 times) the Wherry-Smith formula gives results which are more nearly equally distributed

TABLE IV

Correlations Observed and Theoretical (Based upon Observed Medians)

(N = 37 throughout)

M	Observed medians	Correlation predicted		Error	
		Br.-Sp.	Wherry	Br.-Sp.	Wherry
	(Trait 1)				
1	.344				
5	.752	.724	.682	−.028	−.070
10	.663	.840	.779	.177	.116
15	.702	.887	.807	.185	.105
20	.805	.913	.805	.108	.000
30	.936	.940	.635	.004	−.301
	(Trait 5)				
1	.450				
5	.760	.804	.776	.040	.016
10	.856	.891	.852	.035	−.004
15	.931	.925	.873	−.006	−.058
20	.877	.942	.874	.065	−.003
30	.876	.961	.778	.085	−.098
	(Trait 10)				
1	.363				
5	.433	.740	.701	.307	.268
10	.754	.851	.795	.097	.041
15	.872	.895	.822	.023	−.050
20	.898	.919	.820	.021	−.078
30	.872	.945	.669	.073	−.203
	(All Traits)				
1	.503				
20	.898	.953	.879	.055	−.019
30	.872	.968	.829	.986	−.043

between positive and negative errors (7 to 10), tending to slightly underestimate. The actual distribution of errors can be more easily seen by an inspection of Table II.

Now, if our formula were perfectly correct, we should expect that the errors incurred by its use would be normally distributed about a mean error of zero. The Rietz criteria for normality of distribution were applied to these errors with results as shown in Table III.[4] It can be readily seen that the Wherry correction formula gave much better results than did the uncorrected Brown-Spearman formula when measured by the Rietz criteria.

All of the results in the first three tables are based upon the means of the results obtained by Remmers, since this was the method used in his paper. However, when the number of cases is small, as they were in this study, it is

TABLE V

Error in Predicting Reliability (Based upon Observed Medians)

Error	Brown-Spearman	Wherry
over .210	1	1
.151– .210	2	
.091– .150	3	2
.031– .090	5	1
−.029- .030	6	5
−.089– −.030		5
−.149– −.090		1
−.209– −.150		1
below −.209		1

TABLE VI

Rietz Criteria of Normality Applied to Results from Medians

Criterion	Normal	Brown-Spearman	Wherry
u_1	0	.074	−.018
β_1	0	.497	−.081
β_2	3	1.599	2.284

sometimes preferable to use the median rather than the mean as a basis of calculation, since the median is less affected by extreme cases. The writer has therefore recalculated the problem on the basis of the medians discovered by Remmers, and the results are given in Tables IV, V, and VI. The results were found to differ but little from those based upon the means of the distributions.

If we now assume that the formula (3) has been empirically established and justified, we must still answer a very practical question, namely, "How long shall we make our tests in order to achieve the greatest reliability?" To answer this question we must find the point at which $\bar{R}$ becomes a maximum, with respect to changes in M, assuming r_{11} and N to be constant terms. To find this

point we must find the derivative of equation (3) with respect to M and set the numerator equal to zero, thus, if we write Formula (3) in a slightly more usable form, we have,

$$\bar{R}^2 = \frac{(N-1)M^2 r_{11}^2}{(N-M)(1+2[M-1]r_{11}+[M-1]^2 r_{11}^2)} - \frac{M-1}{N-M}, \tag{3a}$$

whence

$$\frac{d\bar{R}^2}{dM} = \frac{(1+[M-1]r_{11})\{4r_{11}^2 M^2 - (2Nr_{11}^2 + 3r_{11}[1-r_{11}])M + (1-r_{11})^2\}}{(N-M)^2(1-2[M-1]r_{11}+[M-1]^2 r_{11}^2)^2} \tag{4}$$

which causes $\bar{R}$ to reach a maximum or minimum when the numerator is placed

TABLE VII

Showing the value of M which will give a maximum value for $\bar{R}$

(According to the Brown-Spearman-Wherry-Smith formula)

r_{11}						
30	.40	.50	.60	.70	.80	.90
3	4	4	4	5	5	5
8	9	9	10	10	10	10
3	14	14	14	15	15	15
8	19	19	19	20	20	20
3	24	24	24	25	25	25
8	29	29	29	30	30	30
3	34	34	34	35	35	35
8	39	39	39	40	40	40
3	44	44	44	45	45	45
8	49	49	49	50	50	50

equal to zero. Thus, placing the numerator equal to zero and factoring this equation, we find its roots to be

$$M = \frac{\cdot(1-r_{11})}{r_{11}} \tag{5a}$$

$$M = \frac{2Nr_{11} - 3(1-r_{11}) - \sqrt{4N^2 r_{11}^2 - 12Nr_{11}(1-r_{11}) - 7(1-r_{11})^2}}{8r_{11}} \tag{5b}$$

or

$$M = \frac{2Nr_{11} - 3(1-r_{11}) + \sqrt{4N^2 r_{11}^2 - 12Nr_{11}(1-r_{11}) - 7(1-r_{11})^2}}{8r_{11}} \tag{5c}$$

and by substituting actual values of N and r_{11} in the equations, we find that equation (5c) is the root we are seeking (i.e.) the value of M for which $\bar{R}$ becomes a maximum.

It can also be readily seen that the value under the radical approximates a perfect square (lacking 16 units of being that figure) of the quantity outside of the radical, thus approximating this value for large values of N. Thus, when N is large (exceeds 100) we may secure satisfactory approximations to M if we rewrite equation (5c) in the form below

$$M_{(\text{Approximately})} = \frac{N}{2} - \frac{3(1 - r_{11})}{4r_{11}}. \tag{5d}$$

Table VII shows the results of equation (5c) for values between $N = 10$ and $N = 100$ (by increments of 10) for values of r_{11} from .10 to .90 (by increments of .10). The use of the formula does not yield integers, and so the results in the table are recorded to the nearest whole number rather than exactly as given by the formula.

If, in order to test the validity of formula (5c), we apply it to the values in Tables I and IV, we find fairly close agreement. The formula in each case predicts a maximum value for $\bar{R}$ when M lies between 15 and 20, and in the actually lengthened tests R is found to be a maximum when M is 30, 15, 15, 20, 30, 15, 20, and 20, thus being in agreement six times out of eight.

Conclusions

1. The Brown-Spearman formula appears to give results which contain both constant and chance errors.

2. These results can be practically eliminated by applying the Wherry-Smith correction formula to the results obtained by the Brown-Spearman formula.

3. We may find the value of M which will give the greatest value of $\bar{R}$ by substitution in equation (5c) above, and then by substitution of this value in equation (3), find the most probable value of $\bar{R}$ at its maximum point.

4. For large values of N we may secure satisfactory approximations to M by means of the simpler formula (5d).

BIBLIOGRAPHY

1. Hull, Clark: "Memorandum Concerning Factors Influencing the Prediction of Performance," Appendix F, Conference of Individual Psychological Differences, National Research Council, Washington, D. C., 1930.
2. Wherry, Robert J.: "A New Formula for Predicting the Shrinkage of the Multiple Correlation Coefficient," Annals of Mathematical Statistics, November, 1931.
3. Remmers, H. H.: "The Equivalence of Judgments to Test Items in the Sense of the Spearman-Brown Formula," Journal of Educational Psychology, January, 1931.
4. Rietz, H. L.: Mathematical Statistics, Carus Mathematical Monograph, No. 3, Chicago, 1927, pp. 58–59.

Cumberland University
Lebanon, Tennessee

THE LIKELIHOOD TEST OF INDEPENDENCE IN CONTINGENCY TABLES

By S. S. Wilks

J. Neyman and E. S. Pearson[1] have applied the principle of the ratio of likelihoods to the problem of determining criteria for testing various hypotheses about the group frequencies in problems dealing with grouped data. In particular, they have discussed the fundamental χ^2 problem, the test of goodness of fit, the hypothesis that two samples of grouped data are from the same population, and the hypothesis of independence in contingency tables. In their treatment of these problems, these authors have started from the limiting form of the probability of an observed set of frequencies and have shown that approximately each of the appropriate λ's is a function of the minimum value of a corresponding χ_1^2. The distribution of this minimum value is found, from which the significance test is made.

In certain cases the exact values of the λ's are relatively simple functions of the observations which can be as conveniently calculated as the corresponding χ^2's. The purpose of this note is to consider the exact expressions for the λ's and find their asymptotic distributions in large samples for the following hypotheses: (1) that a sample of grouped data is from a population with specified group frequencies (i.e., the fundamental χ^2 problem) ,(2) that several samples of grouped data are from the same population, and (3) that there is independence in a contingency table.

1. **The fundamental χ^2 problem.** Let $p_1, p_2, \cdots p_k$ be the probabilities of the mutually exclusive events $E_1, E_2, \cdots E_k$ respectively. In a sample of N events the probability that $E_1, E_2, \cdots E_k$ will occur $n_1, n_2, \cdots n_k$ times respectively, is given by

$$C = \frac{N!}{n_1!\, n_2! \cdots n_k!} p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}. \tag{1}$$

If we let Ω be the class of all sets of values of the p's such that their sum is unity, there is only one set of p's that maximize C, namely, $p_j = n_j/N$ $(j = 1, 2, \cdots k)$. The maximum of C is

$$C(\Omega \max) = \frac{N!}{n_1!\, n_2! \cdots n_k!} \cdot \frac{n_1^{n_1} n_2^{n_2} \cdots n_k^{n_k}}{N^N}. \tag{2}$$

[1] Biometrika, vol. 20A (1928), pp. 263–294.

The likelihood of the hypothesis that the sample is from a population specified by p's having the values $p_1, p_2, \cdots p_k$ is defined as

$$(3) \qquad \lambda_s = \frac{C}{C(\Omega \max)} = \left(\frac{Np_1}{n_1}\right)^{n_1}\left(\frac{Np_2}{n_2}\right)^{n_2}\cdots\left(\frac{Np_k}{n_k}\right)^{n_k}.$$

λ_s is a quantity which clearly lies between 0 and 1. It will be 1 only when $p_j = n_j/N$ $(j = 1, 2, \cdots k)$, (that is, when the hypothesis is rigorously supported by the sample) and tends to 0 as the sample values n_j/N diverge more and more from the hypothetical values p_j. The problem of making an exact test of significance of an observed value of λ_s would involve the computation of all terms of form (1) the n's of which make λ_s less than the observed value of λ_s. This, of course, is impracticable except perhaps for the binomial case with small values of N. However, if the n's are large we can find an approximate solution. If we let $x_j = \dfrac{n_j - Np_j}{\sqrt{N}}$ then except for terms of order $1/\sqrt{N}$ and higher, the x's are distributed according to the law

$$(4) \qquad \frac{1}{\sqrt{(2\pi)^{k-1}p_1p_2\cdots p_k}}\, e^{-\frac{1}{2}\sum_j \frac{x_j^2}{p_j}},$$

where $\Sigma_j x_j = 0$. Neglecting terms of order $1/\sqrt{N}$ and higher we easily find (using natural logarithms) $-2 \log \lambda_s = \sum_j \dfrac{x_j^2}{2p_j}$. Therefore, if $\theta = -2 \log \lambda_s$, θ is approximately distributed according to the function

$$(5) \qquad \frac{(\frac{1}{2})^{\frac{k-1}{2}}\theta^{\frac{k-3}{2}}}{\Gamma\left(\frac{k-1}{2}\right)}e^{-\frac{\theta}{2}}$$

which is the χ^2 distribution with $k - 1$ degrees of freedom.

Since we have neglected terms of order $1/\sqrt{N}$ in obtaining (4) there is no theoretical reason why χ^2 should be used in preference to $-2 \log \lambda_s$ as the criterion for testing the hypothesis that the sample is from a population specified by $p_1, p_2, \cdots p_k$. Any practical advantage which $-2 \log \lambda_s$ may have will therefore justify its use.

2. The hypothesis that several samples of grouped data are from a common population. Let $p_{i1}, p_{i2}, \cdots p_{is}$ be the probabilities with which the mutually exclusive events $E_{i1}, E_{i2}, \cdots E_{is}$ occur, where $\Sigma_j p_{ij} = 1$ $(i = 1, 2, \cdots r)$. Then in a sample of N_i events the chance that $E_{i1}, E_{i2}, \cdots E_{is}$ will occur $n_{i1}, n_{i2}, \cdots n_{is}$ times respectively is given by an expression similar to (1). The chance of the joint occurrence of the r samples is

$$(6) \qquad \frac{N_1!\,N_2!\cdots N_r!}{n_{11}!\,n_{12}!\cdots n_{rs}!}\,p_{11}^{n_{11}}p_{12}^{n_{12}}\cdots p_{rs}^{n_{rs}}.$$

We are interested in testing the hypothesis that the r samples are from the same population, that is, that the r sets of p's $p_{i1}, p_{i2}, \cdots p_{is}$ $(i = 1, 2, \cdots r)$ are the same. The likelihood criterion λ_c appropriate to this hypothesis is the ratio of the maximum (ω(max)) of (6) subject to the condition that the sets of p's are the same (that is, $p_{ij} = p_j$ say, $i = 1, 2, \cdots r$; $j = 1, 2, \cdots s$) to the maximum (Ω (max)) of (6) without this restriction.

For convenience let the observations be arranged in table form so that n_{ij} is the frequency in the i-th row and j-th column. Let $n_{i\cdot}$ and $n_{\cdot j}$ be the totals of the i-th row and j-th column respectively, and N the total of all observations. Thus $n_{i\cdot}$ is the same as N_i. The expression for λ_c will be

$$\lambda_c = \frac{n_{\cdot 1}^{n_{\cdot 1}} n_{\cdot 2}^{n_{\cdot 2}} \cdots n_{\cdot s}^{n_{\cdot s}} n_{1\cdot}^{n_{1\cdot}} n_{2\cdot}^{n_{2\cdot}} \cdots n_{r\cdot}^{n_{r\cdot}}}{N^N n_{11}^{n_{11}} n_{12}^{n_{12}} \cdots n_{rs}^{n_{rs}}}. \tag{7}$$

It can be shown analytically that λ_c lies between 0 and 1. It can be 1 only when $\frac{n_{1j}}{N_1} = \frac{n_{2j}}{N_2} = \cdots = \frac{n_{rj}}{N_r}$, $j = 1, 2, \cdots s$, that is, when the hypothesis of a common population is perfectly substantiated by the samples. Because of the fact that the n_{ij} are integers, it is clear that λ_c can be 1 only in exceptional cases, but it can take on values arbitrarily near 1 for sufficiently large values of the n_{ij}.

If the N_i are large, the quantities $x_{ij} = \dfrac{n_{ij} - N_i p_j}{\sqrt{N_i}}$ are approximately distributed according to the function

$$F = \left(\frac{1}{(2\pi)^{s-1} p_1 p_2 \cdots p_s}\right)^{\frac{r}{2}} e^{-\frac{1}{2}\sum_{i,j}\frac{x_{ij}^2}{p_j}}, \tag{8}$$

where $\Sigma_j x_{ij} = 0$, $i = 1, 2, \cdots r$. By neglecting terms of order $1/\sqrt{N}$ and higher, we find that

$$-2 \log \lambda_c = \sum_{i,j} \frac{(N x_{ij} - \sqrt{N_i}(\sum_i \sqrt{N_i}\, x_{ij}))^2}{N^2 p_j}. \tag{9}$$

Denoting the quantity on the right side of (9) by χ_0^2 it follows by straightforward analysis that the characteristic function $\varphi(t)$ of χ_0^2 defined by the $r(s - 1)$-tuple integral $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{it\chi_0^2} F\, dx_{11} \cdots dx_{rs}$ has the value

$$(\tfrac{1}{2})^{\frac{(r-1)(s-1)}{2}} (\tfrac{1}{2} - it)^{-\frac{(r-1)(s-1)}{2}}. \tag{10}$$

But it is well known that (10) is the characteristic function of any quantity distributed according to (5) with $(k - 1)$ replaced by $(r - 1)(s - 1)$. This, of course, is the χ^2 distribution with $(r - 1)(s - 1)$ degrees of freedom.

It will be noticed that the exact value of λ_c is a function of the observations n_{ij} which is independent of the p's, while the approximate value of $-2 \log \lambda_c$

as given by (9) involves the p's. Before (9) could be used practically, one would have to replace the p's by sample estimates, thus making further approximations necessary in order to get the distribution. If the usual estimates $p_j = n_{.j}/N$ are used for the p's in χ_0^2 we find that χ_0^2 reduces to

$$(11) \qquad \sum_{i,j} \frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{N}\right)^2}{\frac{n_{i.}n_{.j}}{N}}$$

which is the familiar χ^2 function for testing independence in contingency tables. However, (11) differs from χ_0^2 by terms of the same order (i.e., $1/\sqrt{N_i}$) as those by which χ_0^2 differs from $-2 \log \lambda_c$. Since we have neglected terms of the same order in obtaining (8), there is no theoretical reason why (11) should be used rather than $-2 \log \lambda_c$ for testing the hypothesis that the m samples are from a common population.

3. **The hypothesis of independence in contingency tables.** We shall consider a sample of N observations which can be arranged in a two-way contingency table having r rows and s columns. Let p_{ij} be the probability that an observation will fall in the i-th row and j-th column. The probability that the sample of N items will be distributed so that n_{ij} will be the number falling in the i-th row and j-th column ($i = 1, 2, \cdots r$; $j = 1, 2, \cdots s$) is given by

$$(12) \qquad \frac{N!}{n_{11}!\, n_{12}! \cdots n_{rs}!} p_{11}^{n_{11}} p_{12}^{n_{12}} \cdots p_{rs}^{n_{rs}}.$$

Here we are interested in testing the hypothesis that the classification by rows is independent of the classification by columns, that is, that p_{ij} is of the form $p_i q_j$ where

$$(13) \qquad \Sigma_i p_i = 1, \qquad \Sigma_j q_j = 1.$$

For this hypothesis the appropriate likelihood criterion, say λ_c', is the ratio of the maximum (ω(max)) of (12) when $p_{ij} = p_i q_j$ restricted by the conditions (13) to the maximum (Ω (max)) of (12) subject only to the condition that $\sum_{i,j} p_{ij} = 1$. λ_c' turns out to be identical with λ_c in (7). When the hypothesis of independence is true, the approximate distribution of the quantity $-2 \log \lambda_c'$ is the same as that of $-2 \log \lambda_c$ when the hypothesis of a common population is true. To show that the distributions are the same we note that by placing

$$(14) \qquad x_{ij} = \frac{n_{ij} - Np_i q_j}{\sqrt{N}},$$

we find from (12) that the x_{ij} are approximately distributed according to the function

$$(15) \qquad \frac{1}{(2\pi)^{\frac{}{2}} (p_1 p_2 \cdots p_r)^{\frac{}{2}} (q_1 q_2 \cdots q_s)^{\frac{}{2}}} e^{-\frac{1}{2} \sum_{i,j} \frac{x_{ij}^2}{p_i q_j}}$$

where $\sum x_{ij} = 0$. To the same degree of approximation we find

$$(16) \qquad -2 \log \lambda_c' = \sum_{i,j} \frac{(x_{ij} - p_i \sum_i x_{ij} - q_j \sum_j x_{ij})^2}{p_i q_j} = \chi_0'^2$$

Now the characteristic function of $\chi_0'^2$ can be shown without much difficulty to be identical with that of χ_0^2 as given by (10). The identity of the characteristic functions of $\chi_0'^2$ and χ_0^2 implies the identity of the asymptotic distributions of $-2 \log \lambda_c'$ and $-2 \log \lambda_c$. The problem of testing the hypothesis of a common population in several samples of grouped data is mathematically equivalent to that of testing the hypothesis of independence in contingency tables.

If the usual estimates $p_i = \frac{n_i}{N}$, $q_j = \frac{n_j}{N}$ are used in (16) we find that χ_0^2 becomes the expression given by (11). But (11) differs from $\chi_0'^2$ by terms of order $1/\sqrt{N}$ and higher. Therefore, $-2 \log \lambda_c'$ and (11) can differ from each other only by terms of order $1/\sqrt{N}$ which is the order of approximation involved in getting (15) from (12). Thus, $-2 \log \lambda_c'$ has as much validity as the usual criterion (11) for testing for independence in contingency tables.

The λ_c' method can easily be extended to the case of contingency tables of higher order. For example, in a three-way table of r rows, s columns and t layers in which n_{ijk} is the number of items observed in the i-th row, j-th column and k-th layer, the λ_c' criterion for testing the hypothesis of independence, that is, that the probabilities p_{ijk} are of the form $p_{1i} p_{2j} p_{3k}$ is such that

$$(17) \qquad \begin{aligned} -2 \log \lambda_c' = 2 \sum_{i,j,k} (n_{ijk} \log n_{ijk}) + 4 N \log N - 2 \sum_i (n_{i..} \log n_{i..}) \\ - 2 \textstyle\sum_j (n_{.j.} \log n_{.j.}) - 2 \sum_k (n_{..k} \log n_{..k}) \end{aligned}$$

where $n_{i..} = \sum_{j \cdot k} n_{ijk}$, and so on. $-2 \log \lambda'$ in this case is approximately distributed like χ^2 with $rst - r - s - t + 2$ degrees of freedom.

4. Illustrative examples. To illustrate the use of λ_s we shall consider the following example given by R. A. Fisher[2] dealing with de Winton and Bateson's data on results of interbreeding the hybrid (F_1) generation of *Primula* in which two factors are considered.

	Flat Leaves		Crimped Leaves		Total
	Normal Eye	Primrose Queen Eye	Lee's Eye	Primrose Queen Eye	
Observed (n_i)................	328	122	77	33	560
Expected (Np_i)..............	315	105	105	35	560

[2] Statistical Methods for Research Workers, 4th ed. p. 84.

If the two factors are Mendelian, that is, segregate independently, the four classes of offspring resulting from interbreeding the F_1 generation are expected to appear in the ratio 9:3:3:1 (assuming all classes equally viable). We wish to test the hypothesis of a 9:3:3:1 ratio. It is found that

$$-2 \log_e \lambda_s = 2 \log_e 10 \left[\sum_i n_i \log_{10} n_i - \sum_i n_i \log_{10} (Np_i)\right] = 11.50 .$$

Entering Fisher's χ^2 table for $n = 3$, we find that the chance of exceeding the value 11.50 is less than .01, which is significant if we take $P = .05$ as the critical level of significant deviation. Thus, the observed frequencies cannot be reasonably explained as chance deviations from the 9:3:3:1 ratio.

The usual χ^2 method gives $\chi^2 = 10.87$ and $n = 3$ for the 9:3:3:1 hypothesis. The value of P in this case lies between .01 and .02. It follows from the theoretical discussion that 10.87 has no greater validity than 11.50 in testing this hypothesis.

We shall illustrate the use of λ_c by using another example given by Fisher dealing with Wachter's data for back-crosses in mice.

	Black Self	Black Piebald	Brown Self	Brown Piebald	Total
Coupling:					
F_1 Males	88	82	75	60	305
F_1 Females	38	34	30	21	123
Repulsion:					
F_1 Males	115	93	80	130	418
F_1 Females	96	88	95	79	358
Total	337	297	280	290	1204

The back-crosses were made according as the male or female parents of the F_1 generation were heterozygous in the two factors Black-Brown, Self-Piebald, and according to whether the two dominant genes came both from one parent (Coupling) or one from each parent (Repulsion). We wish to test the hypothesis that the proportions are independent of the matings used. We find

$$-2 \log \lambda_c = 2 \log_e 10 \left[\sum n_{ij} \log_{10} n_{ij} + N \log_{10} N - \sum_i n_{i.} \log_{10} n_{i.} - \sum_j n_{.j} \log_{10} n_{.j}\right] = 21.69$$

Entering Fisher's χ^2 table for $n = 9$ we find that the chance of exceeding this value is less than .01. The departure from the hypothesis of independence is significant on basis of the $P = .05$ level. The χ^2 method gives the remarkably close result $\chi^2 = 21.83$, which, with $n = 9$ gives $P < .01$.

5. **Summary.** We have considered the exact expressions for the Neyman-Pearson λ criteria appropriate to the following hypotheses: (1) That a sample

of grouped data is from a population with specified group proportions (the fundamental χ^2 problem), (2) that several samples of grouped data are from a common population, (3) that there is independence in a contingency table. The quantity $-2 \log \lambda$ for each of these cases is approximately distributed like χ^2, the number of degrees of freedom being given in each case. It is shown that the usual χ^2 method of testing these hypotheses has no greater theoretical validity than the λ method. On the practical side, it is to be remarked that $-2 \log \lambda$ can be computed with fewer operations than χ^2. Two examples are given to illustrate the practical application of the λ method.

PRINCETON UNIVERSITY.

THE PROBABILITY THAT THE MEAN OF A SECOND SAMPLE WILL DIFFER FROM THE MEAN OF A FIRST SAMPLE BY LESS THAN A CERTAIN MULTIPLE OF THE STANDARD DEVIATION OF THE FIRST SAMPLE

By G. A. Baker, Ph.D.

The following statement of the significance of a probable error is often made: "The probable error of the mean is a value above and below the mean such that if the test were repeated under the same conditions there would be, on the average, equal chances that the mean would fall within or without this range." The probable error is attached to the mean of the sample and it is assumed that the standard deviation of the sample is that of the sampled normal population. This was formerly a very usual explanation of the meaning of probable error by research workers, but it is inaccurate and misleading, especially for samples of 20 or less such as are dealt with in agricultural experiments. The inaccuracy of this explanation of the meaning of probable error has been realized for many years by competent statisticians, but no satisfactory treatment has heretofore been devised.[1]

The attempted explanation of the probable error in terms of the expected frequency of the occurrence of different size deviations of the means of future samples from the sample mean does raise a very interesting, important, and legitimate question, namely, what is the probability of a second mean lying within a certain multiple of the standard deviation of a first sample of the mean of a first sample? This question is of fundamental concern to those engaged in experimental work. Its answer will indicate to investigators reasonable deviations from the results of their first experiments, will form a valid basis for the rejection of doubtful observations or groups of such observations, and will form a basis for a test of the significance of the divergence of results in different experiments. It is found that the usual method of treating the probable error gives an overly optimistic idea of the smallness of the deviations that may be expected in future samples.

The distribution function of the variable

$$v = \frac{\quad}{y}$$

where x is the mean of the first sample, z is the mean of the second sample, and y is the standard deviation of the first sample, is obtained in this paper. The sampled population is assumed to be normal.

[1] Camp, Burton H. "Suggested Problems for Mathematical Research," *Journal American Statistical Association*, Supplement Vol. 30, No. 189A, Mar. 1935, p. 259, No. 5.

Let the sampled population be represented by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad -\infty \leqq x \leqq \infty. \tag{1}$$

If a sample of n is drawn from (1) the means, as is well known, will be distributed as proportional to

$$e^{-\frac{1}{2}nx^2}, \qquad -\infty \leqq x \leqq \infty, \tag{2}$$

and the standard deviation will be distributed as proportional to

$$y^{n-2}e^{-\frac{1}{2}ny^2}, \qquad 0 \leqq y \leqq \infty. \tag{3}$$

If a second sample of n is drawn from (1) its mean will be distributed as proportional to

$$e^{-\frac{1}{2}nz^2}, \qquad -\infty \leqq z \leqq \infty. \tag{4}$$

Consider the expression

$$\frac{x - z}{y} \tag{5}$$

and call it v. Then v is the difference between the means of the two samples measured in terms of the standard deviation of the first sample. The distribution function of v is sought.

The three variables x, y, and z are independent. Let y, for the moment, have a constant value and write

$$vy = x - z. \tag{6}$$

The probability of a given value of vy in $d(vy)$ for a given value of y is now being sought, that is, vy is regarded as constant. This probability is proportional to

$$\left[y^{n-2}e^{-\frac{1}{2}ny^2}e^{-\frac{1}{4}nv^2y^2}\int_{-\infty}^{\infty} e^{-(z+\frac{1}{2}vy)^2}\, dz\right] d(vy), \tag{7}$$

from the application of the following

Lemma. If x and y are independent variables, $-\infty \leqq x \leqq \infty$, $-\infty \leqq y \leqq \infty$, and the probability of an x in dx is $f(x)dx$ and the probability of a y in dy is $\varphi(y)dy$, then the probability of $v = y - x$ in dv is proportional to[2]

$$\left[\int_{-\infty}^{\infty} f(x)\,\varphi(v + x)\, dx\right] dv.$$

Thus the probability of a value of v in dv for a given y is proportional to

$$y^{n-1}e^{-\frac{1}{2}n\left[1+\frac{v^2}{2}\right]y^2}\, dv \tag{8}$$

[2] Baker, G. A. "Random Sampling from Non-Homogeneous Populations," *Metron*, Vol. 8, No. 3, Feb. 1930, p. 68 (slightly modified).

since $d(vy) = ydv$ for y constant. Hence the total probability of a particular value of v in dv will be given as proportional to

$$(9) \qquad \left[\int_0^\infty y^{n-1} e^{-\frac{1}{2}n\left(1+\frac{v^2}{2}\right)v^2}\, dy\right] dv$$

which is proportional to

$$(10) \qquad \frac{dv}{\left(1+\frac{v^2}{2}\right)^{\frac{n}{2}}}$$

If the number in the first sample is n_1 and the number in the second sample is n_2, then (10) becomes

$$(11) \qquad \frac{dv}{\left(1+\frac{n_2}{n_1+n_2}v^2\right)^{\frac{n_1}{2}}}.$$

This distribution, (11), permits an answer to be given to the question, what is the probability that the mean of a sample of a given size n_2 will differ from the mean of a first sample of size n_1 by as much as a constant multiple of the standard deviation of the first sample? Thus, this distribution gives a clear and comprehensible indication of the expected conformity of future experiments and gives a valuable test for the significance of the difference between two means. If it is desired to use this distribution as a rejection criterion, n_1 should be taken so as to include as many items as possible and so as to exclude the doubtful ones. The doubtful items should be included in the second sample. If the original sample is broken up into two or more samples it must be done in such a way as not to destroy the randomness of the resulting parts.

Example. Suppose for the purpose of illustration that a sample of four is to be considered. The proper v-distribution is

$$\frac{\sqrt{2}\, dv}{\pi\left(1+\frac{v^2}{2}\right)^2}$$

The value of v which is necessary to give a probability of one-half is a root of

$$\tan^{-1}\frac{\rho}{\sqrt{2}} + \frac{1}{2}\frac{\sqrt{2}\,\rho}{\rho^2+2} - \frac{\pi}{4}$$

which is .9. That is, an interval of 1.8 times the standard deviation of the sample of four with center at the mean of the sample is necessary for a probability of one-half that the mean of the next sample of four will lie in this interval. This compares with .75 times the standard deviation of the sample if

$$\frac{\sigma}{\sqrt{n-1}}$$

is used as the probable error of the mean and with .65 times the standard deviation of the sample if

$$\frac{\sigma}{\sqrt{n}}$$

is used as the probable error of the mean. The last two methods of calculating a probable error with the interpretation indicated at the beginning of this paper give the investigator an unwarranted feeling of assurance about the agreement of future samples with a first sample.

If two samples of n_1 and n_2 are drawn from the normal population, (1), then these samples can be combined for the purpose of calculating a standard deviation and the difference between the means of the samples can be measured in terms of the standard deviation of the combined sample. The distribution function of the difference of the means divided by the standard deviation of the combined sample is

$$\frac{dv}{\left[1 + \frac{n_1 n_2}{(n_1 + n_2)^2}\, v^2\right]^{\frac{n_1+n_2}{2}}} \tag{11'}$$

This distribution, (11′), is the basis for a valid test for the significance of the difference between two means. If either this test or the test based on distribution (11) shows a significant difference between the means it can not be ignored.

"Student's" t-distribution is proportional to

$$\frac{dt}{\left(1 + \frac{t^2}{N-1}\right)^{\frac{N}{2}}} \tag{12}$$

The above distributions can be easily transformed into t-distributions so that "Student's" tables can be used. For instance, if we put

$$v = \frac{\sqrt{2}\, t}{\sqrt{n-1}}, \qquad N = n,$$

then (10) becomes proportional to (12). Again, put

$$v = \frac{\sqrt{n_1 + n_2}\, t}{\sqrt{n_2}\sqrt{n_1 - 1}}, \qquad N = n_1,$$

and (11) becomes proportional to (12). Finally, put

$$v = \frac{(n_1 + n_2)\, t}{\sqrt{n_1 n_2}\sqrt{n_1 + n_2 - 1}}, \qquad N = n_1 + n_2,$$

and (11′) becomes proportional to (12).

Summary. The distributions found for the difference of the means of two samples in terms of a standard deviation of one sample or combination of both

samples are similar to and easily transformed into "Student's" t-distribution so that his tables can be used. However, these distributions answer a practical, interesting, and important question that "Student's" t-distribution does not. If in an experimental science a series of observations is made it is desirable to know how much a similar series of observations could be expected to differ from the set of observations now available. This deviation, if it is to mean anything, must be expressed in terms of quantities available from the observations already made. This paper gives the probability function of a deviation in the mean of a future sample measured from the mean of a first sample and measured in terms of the standard deviation of a first sample, that is, in terms of quantities known from the first sample. It is a very definite advantage and a great gain in assurance to know the point from which measurements are being made and the unit in which they are expressed instead of making vague, ill-defined assumptions about the zero point and unit length of the measuring scale. It is true that differences that were formerly considered significant may not be so considered now. But these differences would appear insignificant if experiments were sufficiently repeated, so that the net result is fewer inconsistencies to explain away.

ON SAMPLES FROM A MULTIVARIATE NORMAL POPULATION[1]

By Solomon Kullback

1. **Introduction.** In this paper we shall discuss the distribution of certain functions calculated for samples drawn from a multivariate normal population. The method of solution is based on the theory of characteristic functions and presents further application of that theory to the distribution problem of statistics.[2]

We shall have occasion to refer to the multivariate normal population whose distribution law is given by

$$F(x) \equiv \pi^{-n/2} \mid B_{pq} \mid^{1/2} e^{-B(x-m,\, x-m)} \qquad (p, q = 1, 2, \cdots, n) \tag{1.1}$$

where $B(x - m, x - m)$ is the real, positive definite quadratic form of the $x_p - m_p$ with matrix $\| B_{pq} \|$. Here m_p is the mean in the population of the pth variate and $B_{pq} = \Delta_{pq}/2\sigma_p\sigma_q\Delta$ where σ_p is the standard deviation in the population of the pth variate; Δ is the determinant of population correlations $\rho_{pq} = \rho_{qp}$; Δ_{pq} is the co-factor of ρ_{pq} in Δ; and $\mid B_{pq} \mid$ is the determinant of the matrix $\| B_{pq} \|$.

Since the integral of (1.1) over the entire field of variation of the variables is unity, we have (using abbreviated notation)

$$\int e^{-B(x-m,\, x-m)}\, dx = \pi^{n/2} \mid B_{pq} \mid^{-1/2} \tag{1.2}$$

Equation (1.2) will be true if $\| B_{pq} \|$ is complex, provided its real part is symmetric and positive definite.[3]

The distribution of sample means of samples from the population (1.1) is independent of the distribution of the system of sample variances and covariances and is given by[4]

$$F_1(\bar{x}) \equiv \pi^{-n/2} \mid A_{pq} \mid^{1/2} e^{-A(\bar{x}-m,\, \bar{x}-m)} \tag{1.3}$$

where $A(\bar{x} - m, \bar{x} - m)$ is the real, positive definite quadratic form of the $\bar{x}_p - m_p$ with matrix $\| A_{pq} \|$. Here $\bar{x}_p = (1/N) \sum^{N} x_{pa}$ is the sample mean of the pth

[1] Presented to the American Mathematical Society, February 23, 1935.

[2] For more complete reference to the theory of characteristic functions as applied to statistics see S. Kullback, *Annals of Mathematical Statistics*, Vol. 5 (1934), pp. 263–307.

[3] J. Wishart and M. S. Bartlett, *Proc. Cambridge Phil. Soc.*, Vol. 29 (1933), pp. 260 ff.

[4] J. Wishart, *Biometrika*, Vol. 20 A (1928), pp. 32–52.

J. Wishart and M. S. Bartlett, *loc. cit.*

variate, and $A_{pq} = NB_{pq}$, where B_{pq} has been defined for equation (1.1). The distribution law of the system of sample variances and covariances is given by[5]

$$(1.4) \qquad F_2(a) \equiv \frac{|A_{pq}|^{(N-1)/2}}{\pi^{n(n-1)/4} \prod^{n} \Gamma(N-r)/2} e^{-A(a)} \quad {}_{pq}\,|$$

where $A(a) = \sum_{p,q=1}^{n} A_{pq}\, a_{pq}$ and $a_{pq} = a_{qp} = (1/N) \sum_{\alpha=1}^{N} (x_{p\alpha} - \bar{x}_p)(x_{q\alpha} - \bar{x}_q)$ with A_{pq} and $\bar{x}_p$ defined as for (1.3). Since the integral of (1.4) over the entire field of variation of the a_{pq} is unity, we have[6]

$$(1.5) \qquad \int e^{-A(a)} \,|a_{pq}|^{(N-n-2)/2}\, da = \pi^{n(n-1)/4}\, |A_{pq}|^{(1-N)/2} \prod_{r=1}^{n} \Gamma(N-r)/2$$

Equation (1.5) will also hold if the matrix $\| A_{pq} \|$ is complex, provided its real part is symmetric and positive definite.[7]

2. **Variance.** Consider a sample of N independent items from the normal population (1.1). Let

$$(2.1) \qquad v = \sum^{n} a_{p}$$

where a_{pq} is defined as in (1.4). From the theory of characteristic functions and (1.5), we have that the characteristic function of the distribution law of v is given by[8]

$$(2.2) \qquad \varphi(t) = \int e^{it\Sigma a_{pq}} F_2(a)\, da = |A_{pq}|^{(N-1)/2} \cdot |A_{pq} - it|^{(1-N)/2}.$$

It may be readily shown that

$$(2.3) \qquad |A_{pq} - it| = |A_{pq}| - it \sum_{p,q=1}^{n} A^{pq}$$

where A^{pq} is the co-factor of A_{pq} in $|A_{pq}|$.

We thus have for the distribution law[8] of v

$$(2.4) \qquad P(v) = (A/c)^{(N-1)/2} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itv} (A/c - it)^{(1-N)/2}\, dt$$

[5] J. Wishart, *loc. cit.*

[6] Cf. S. S. Wilks, *Biometrika*, Vol. 24 (1932), pp. 471-494.

[7] A. E. Ingham, *Proc. Cambridge Phil. Soc.*, Vol. 29 (1933), p. 271 ff. The considerations in this paper will still hold if the condition above is imposed.

[8] S. Kullback, *loc. cit.*, p. 272.

where $A = |A_{pq}|$, $c = \sum_{p,q=1}^{n} A^{pq}$ and $A/c > 0$ since $\| A_{pq} \|$ is positive definite. By using the fact that[9]

$$(2.5) \qquad \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} e^{\mu z} z^{-k} \, dz = \begin{cases} \mu^{k-1}/\Gamma(k), & \mu > 0 \\ 0, & \mu \leqq 0 \end{cases}$$

where $k > 0$, $a > 0$, we have

$$(2.6) \qquad P(v) = \frac{(A/c)^{(N-1)/2}}{\Gamma(N-1)/2} v^{(N-3)/2} e^{-(A/c)v}$$

3. **Ratio of variances.** If v_1 and v_2 represent the statistic v (defined in (2.1)), obtained from independent samples of N_1 and N_2 items respectively, then it may be shown that the distribution law of $w = v_1/v_2$ is given by[10]

$$(3.1) \qquad P(w) = \frac{\Gamma(N_1 + N_2 - 2)/2}{\Gamma(N_1 - 1)/2\, \Gamma(N_2 - 1)/2} w^{(N_1-3)/2}(1 + w)^{(2-N_1-N_2)/2}.$$

If we set $w = e^{2z} n_1/n_2$, where $n_1 = N_1 - 1$ and $n_2 = N_2 - 1$ we obtain for the distribution law of z[11]

$$(3.2) \qquad P(z) = 2 \frac{\Gamma(n_1 + n_2)/2}{\Gamma n_1/2\ \Gamma n_2/2} n_1^{n_1/2} n_2^{n_2/2} e^{n_1 z}(n_2 + n_1 e^{2z})^{-(n_1+n_2)/2}.$$

4. **Student's distribution.** Consider a sample of N independent items from the normal population (1.1). Let

$$(4.1) \qquad \mu = \sum_{p,q=1}^{n} (\bar{x}_p - m_p)(\bar{x}_q - m_q)$$

where $\bar{x}_p$ and m_p are defined as in (1.3). The characteristic function of the simultaneous distribution function of μ, defined as in (4.1) and v defined as in (2.1) is given by

$$(4.2) \qquad \varphi(t_1, t_2) = \int \exp\left\{ it_1 \sum_{p,q=1}^{n} (\bar{x}_p - m_p)(\bar{x}_q - m_q) + it_2 \sum_{p,q=1}^{n} a_{pq} \right\} F_1(\bar{x})F_2(a)\, d\bar{x}da$$

[9] Cf. A. E. Ingham, *loc. cit.*
J. Wishart and M. S. Bartlett, *Proc. Cambridge Phil. Soc.*, Vol. 28 (1932), p. 455 ff.
[10] S. Kullback, note accepted for publication soon in the *Annals of Math. Statistics.*
[11] Cf. R. A. Fisher, I. *Proc. International Math. Congress, Toronto* (1924), Vol. 2, pp. 805–813.
R. A. Fisher, II. *Statistical Methods for Research Workers*, 4th Edition (1932), Edinburgh: Oliver and Boyd, pp. 224–227.

where F_1 and F_2 are defined as in (1.3) and (1.4) respectively. From (1.2) and (1.5) we have that

$$\varphi(t_1, t_2) = (A/c)^{N/2}(A/c - it_1)^{-1/2}(A/c - it_2)^{(1-N)/2} \tag{4.3}$$

where A and c are defined as in (2.4). The simultaneous distribution of μ and v is given by

$$P(\mu, v) = (1/2\pi)^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-it_1\mu - it_2 v}\varphi(t_1, t_2)\, dt_1 dt_2 \tag{4.4}$$

which evaluated by a procedure similar to that used for (2.4) yields

$$P(\mu, v) = \frac{(A/c)^{N/2}}{\Gamma(N-1)/2\,\Gamma 1/2}\,\mu^{-1/2}e^{-\mu A/c}v^{(N-3)/2}e^{-vA/c}. \tag{4.5}$$

From (4.5) we may readily obtain the distribution of $z = \mu^{1/2}/v^{1/2}$ to be[12]

$$D(z) = 2\,\frac{\Gamma N/2}{\Gamma(N-1)/2\,\Gamma 1/2}\,(1 + z^2)^{-N/2}, \qquad (0 \leqq z \leqq \infty). \tag{4.6}$$

5. **k samples.** Suppose we have k independent samples of $N_1, N_2, \cdots, N_k$ items respectively, drawn fron the normal population defined by (1.1). Let μ_r, $(r = 1, 2, \cdots, k)$ be the statistic μ, defined by (4.1), for each of the k samples respectively; let V_r, $(r = 1, 2, \cdots, k)$ be the statistic V, defined by (2.1), for each of the k samples respectively; let μ_0 and V_0 be the values of these statistics for the sample of $N = N_1 + N_2 + \cdots + N_k$ items obtained by pooling all the samples.

It may be readily verified that

$$\mu_0 = \sum_{r=1}^{k} \mu_r N_r^2/N^2 + 2\sum_{\alpha,\beta=1}^{k} \mu_\alpha^{1/2}\mu_\beta^{1/2}N_\alpha N_\beta/N^2 \qquad (\alpha \neq \beta) \tag{5.1}$$

$$N\mu_0 + NV_0 = \sum_{r=1}^{k}(N_r\mu_r + N_rV_r) \tag{5.2}$$

$$NV_0 = \sum_{r=1}^{k}(N_rV_r + M_r\mu_r) - 2\sum_{\alpha,\beta=1}^{k}\mu_\alpha^{1/2}\mu_\beta^{1/2}N_\alpha N_\beta/N \qquad (\alpha \neq \beta) \tag{5.3}$$

where $M_r = (NN_r - N_r^2)/N$.

In view of (2.6) and (4.5), it is evident that the simultaneous distribution law of μ_r, V_r, $(r = 1, 2, \cdots, k)$ is given by

$$P(\mu)\cdot Q(v) \equiv \prod^{k} P(\mu_r; N_r)\, Q(V_r; N_r) \tag{5.4}$$

[12] Cf. "Student," *Biometrika*, Vol. 6 (1908–09), pp. 1–25.
R. A. Fisher, *Metron*, Vol. 5 (1925), pp. 90–104.
P. R. Rider, *Annals of Mathematics*, 2nd S., Vol. 31 (1930), pp. 579–582.

where

$$(5.5) \qquad P(\mu_r; N_r) \equiv \frac{N_r^{1/2}}{\Gamma 1/2} (B/D)^{1/2} \mu_r^{-1/2} e^{-N_r \mu_r B/D}$$

$$(5.6) \qquad Q(V_r; N_r) \equiv \frac{N_r^{(N_r-1)/2}}{\Gamma(N_r - 1)/2}(B/D)^{(N_r-1)/2} V_r^{(N_r-3)/2} e^{-N_r V_r B/D}$$

and B is the determinant $| B_{pq} |$ defined in (1.1) and $D = \sum_{p,q=1} B^{pq}$ where B^{pq} is the co-factor of B_{pq} in $| B_{pq} |$.

Using (5.3) and (5.4), we find that the characteristic function of the simultaneous distribution law of $\varphi_r = V_r B/D$, $(r = 0, 1, \cdots, k)$ is given by

$$(5.7) \qquad \varphi(t_0, t_1, \cdots, t_k) = \int e^{U(t_0)+V(t_0, t_1, \cdots, t_k)} P(u) \cdot Q(v) du dv$$

where

$$U(t_0) = (B\, it_0/D) \left\{ \sum_{r=1}^{k} \mu_r M_r/N - 2 \sum_{\alpha, \beta=1}^{k} \mu_\alpha^{1/2} \mu_\beta^{1/2} N_\alpha N_\beta/N^2 \right\}, \quad (\alpha \neq \beta)$$

and

$$V(t_0, t_1, \cdots, t_k) = (B/D) \left\{ \sum_{r=1}^{k} V_r(it_r + it_0\, N_r/N) \right\}$$

Let $\mu_r B/D = \zeta_r^2$ and $V_r B/D = \eta_r$, $(r = 1, 2, \cdots, k)$ and rewrite (5.7) as the product of $k + 1$ integrals

$$(5.8) \qquad \varphi(t_0, t_1, \cdots, t_k) = I_0 I_1 \cdots I_k$$

where

$$(5.9) \qquad I_0 = \frac{(N_1 N_2 \cdots N_k)^{1/2}}{\Gamma(1/2)^k} \int e^{-T(\zeta, \zeta)}\, d\zeta$$

with

$$T(\zeta, \zeta) = \sum_{r=1}^{k} \zeta_r^2 (N_r - it_0\, M_r/N) + 2\, it_0 \sum^{k} \zeta_\alpha \zeta_\beta N_\alpha N_\beta/N^2, \quad (\alpha \neq \beta)$$

and

$$(5.10) \quad I_r = \frac{N_r^{(N_r-1)/2}}{\Gamma(N_r - 1)/2} \int_0^\infty \exp\{- \eta_r\, (N_r - it_0 N_r/N - it_r)\}\, \eta_r^{(N_r-3)/2}\, d\eta_r\,.$$

By employing (1.2) we find that

$$(5.11)\quad I_0 = (N_1N_2\cdots N_k)^{1/2}\begin{vmatrix} N_1 - it_0M_1/N & it_0\,N_1N_2/N^2 & \cdots & it_0\,N_1N_k/N^2 \\ it_0\,N_2N_1/N^2 & N_2 - it_0\,M_2/N & \cdots & it_0\,N_2N_k/N^2 \\ \cdots & \cdots & \cdots & \cdots \\ it_0\,N_kN_1/N^2 & it_0\,N_kN_2/N^2 & \cdots & N_k - it_0\,M_k/N \end{vmatrix}^{-1/2}$$

The determinant may be readily evaluated by removing the common factor N_r from the rth row (remembering the value of M_r as given in (5.3)) and applying the operations[13] (row 1 − row 2), (row 2 − row 3), $\cdots$, and then column k + column 1 + column 2 + $\cdots$ + column $k - 1$. We thus obtain

$$(5.12)\qquad I_0 = (1 - it_0/N)^{-(k-1)/2}$$

The integral in (5.10) is well-known and yields

$$(5.13)\qquad I_r = N_r^{(N_r-1)/2}\,(N_r - it_0\,N_r/N - it_r)^{-(N_r-1)/2}.$$

There thus results

$$(5.14)\quad \varphi(t_0, t_1, \cdots, t_k) = G(1 - it_0/N)^{-(k-1)/2}\prod^{k}(N_\alpha - it_0\,N_\alpha/N - it_\alpha)^{-(N_\alpha-1)/2}$$

where $G = \prod_{\alpha=1}^{k} N_\alpha^{(N_\alpha-1)/2}$.

The simultaneous distribution law of φ_r, $(r = 0, 1, \cdots, k)$ is given by

$$(5.15)\quad P(\varphi_0, \varphi_1, \cdots, \varphi_k) = \frac{G}{(2\pi)^{k+1}}\int_{-\infty}\frac{e^{-it_0\varphi_0-it_1\varphi_1-\cdots-itk\,\varphi k}\,dt_0\,dt_1\cdots dt_k}{(1 - it_0/N)^{(k-1)/2}\prod_{\alpha=1}^{k}(N_\alpha - it_0\,N_\alpha/N - it_\alpha)^{(N_\alpha-1)/2}}$$

Integrating successively with respect to $t_k, t_{k-1}, \cdots, t_1$ and applying (2.5) we have

$$(5.16)\quad P(\varphi_0, \varphi_1, \cdots, \varphi_k) = G\exp\left\{-\sum^{k} N_\alpha\varphi_\alpha\right\}\prod_{\alpha=1}\frac{\varphi_\alpha^{(N_\alpha-3)/2}}{\Gamma(N_\alpha-1)/2}\,\frac{1}{2\pi}\int\frac{e^{-it_0\left(\varphi_0-\frac{N_1}{N}\varphi_1-\cdots-\frac{N_k}{N}\varphi_k\right)}\,dt_0}{(1 - it_0/N)^{(k-1)/2}}$$

13 Cf. A. C. Aitken, *Quarterly Journal Math.*, Vol. 2 (1931), pp. 130–135.

and finally

$$P(\varphi_0, \varphi_1, \cdots, \varphi_k) = GN^{(k-1)/2}e^{-N\varphi_0}$$

(5.17)
$$\frac{(\varphi_0 - N_1\varphi_1/N - \cdots - N_k\varphi_k/N)^{(k-3)/2}}{\Gamma(k-1)/2} \prod_{\alpha=1}^{k} \frac{\varphi_\alpha^{(N_\alpha-3)/2}}{\Gamma(N_\alpha - 1/2)}.$$

If we apply to (5.17) the transformation

(5.18)
$$\begin{cases} \varphi_0 = \varphi_0 \\ \varphi_r = N\zeta_r\varphi_0/N_r \qquad (r = 1, 2, \cdots, k) \end{cases}$$

and integrate out φ_0, we obtain for the simultaneous distribution law of $\zeta_r = N_r\varphi_r/N\varphi_0 = N_rV_r/NV_0$

(5.19)
$$D(\zeta_1, \zeta_2, \cdots, \zeta_k) = \frac{\Gamma(N-1)/2}{\Gamma(k-1)/2}(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_k)^{(k-3)/2} \prod_{\alpha=1}^{k} \frac{\zeta_\alpha^{(N_\alpha-3)/2}}{\Gamma(N_\alpha - 1)/2}$$

where the limits of variation in (5.19) are[14]

(5.20)
$$\begin{cases} 0 \leqq \zeta_1 \leqq 1 \\ 0 \leqq \zeta_r \leqq 1 - \zeta_1 - \zeta_2 - \cdots - \zeta_{r-1}, \qquad (r = 2, 3, \cdots, k) \end{cases}$$

6. **Correlation ratio.** Let $\zeta = \log(1 - \zeta_1 - \zeta_2 - \cdots - \zeta_k)$ where the ζ_r, $(r = 1, 2, \cdots, k)$ are defined and distributed as in (5.19). The characteristic function of the distribution law of ζ is given by

(6.1)
$$\varphi(t) = \frac{\Gamma(N-1)/2}{\Gamma(k-1)/2}\int (1 - \zeta_1 - \zeta_2 - \cdots - \zeta_k)^{(k+2it-3)/2} \prod_{\alpha=1}^{k} \frac{\zeta_\alpha^{(N_\alpha-3)/2}}{\Gamma(N_\alpha - 1)/2} d\zeta_\alpha$$

where the limits of variation are given by (5.20). The integral in (6.1) is readily evaluated as a Dirichlet integral,[15] and we obtain

(6.2)
$$\varphi(t) = \frac{\Gamma(N-1)/2\ \Gamma(k - 1 + 2it)/2}{\Gamma(k-1)/2\ \Gamma(N - 1 + 2it)/2}.$$

[14] Cf. J. Neyman and E. S. Pearson, I. *Bulletin de l'Académie Polonaise des Sciences et des Lettres, Série A, Sciences Mathématiques*, 1931, pp. 460–481.

[15] E. Goursat-E. R. Hedrick, *Mathematical Analysis*, Vol. I (1904) (Ginn and Co., N. Y.), p. 308.

The distribution law of ζ is given by

$$(6.3)\qquad P(\zeta) = \frac{\Gamma(N-1)/2}{\Gamma(k-1)/2}\cdot\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it\zeta}\,\frac{\Gamma(k-1+2it)/2}{\Gamma(N-1+2it)/2}\,dt\,.$$

Now it may be shown that[16]

$$(6.4)\qquad \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-it\zeta}\,\frac{\Gamma(k-1+2it)/2}{\Gamma(N-1+2it)/2}\,dt = \frac{e^{\zeta(k-1)/2}(1-e^{\zeta})^{(N-k-2)/2}}{\Gamma(N-k)/2}$$

so that

$$(6.5)\qquad P(\zeta) = \frac{\Gamma(N-1)/2}{\Gamma(k-1)/2\,\Gamma(N-k)/2}\,e^{\zeta(k-1)/2}(1-e^{\zeta})^{(N-k-2)/2}.$$

If we set $e^{\zeta} = \eta^2$, then we obtain for the distribution[17] of η^2

$$(6.6)\qquad D(\eta^2) = \frac{\Gamma(N-1)/2}{\Gamma(k-1)/2\,\Gamma(N-k)/2}\,(\eta^2)^{(k-3)/2}(1-\eta^2)^{(N-k-2)/2}.$$

From its definition we have that

$$(6.7)\qquad \eta^2 = (NV_0 - N_1V_1 - \cdots - N_kV_k)/NV_0$$

which reduces to

$$(6.8)\qquad \eta^2 = (N_1W_1 + N_2W_2 + \cdots + N_kW_k)/NV_0$$

where $W_\alpha = \sum_{p,q=1}^{n} (\bar{x}_{p\alpha} - \bar{x}_{p0})(\bar{x}_{q\alpha} - \bar{x}_{q0})$ with $\bar{x}_{p\alpha}$ the sample mean of the *pth* variate in the *αth* sample and $\bar{x}_{p0}$ the sample mean of the *pth* variate in the sample formed by pooling all the samples.[18]

In a similar manner, we have that the distribution law of $\eta_\alpha^2 = \zeta_\alpha$, $(\alpha = 1, 2, \cdots, k)$ is given by

$$(6.9)\qquad D(\eta_\alpha^2) = \frac{\Gamma(N-1/2)}{\Gamma(N_\alpha-1)/2\,\Gamma(N-N_\alpha/2)}\,(\eta_\alpha^2)^{(N_\alpha-3)/2}(1-\eta_\alpha^2)^{(N-N_\alpha-2)/2}.$$

It may be of interest to point out another derivation for the distribution of $h^2 = 1 - \eta^2$. Let

$$(6.10)\qquad \begin{cases} \theta = (B/D)(N_1V_1 + N_2V_2 + \cdots + N_kV_k) \\ \theta_0 = (B/D)NV_0 \end{cases}$$

[16] Whittaker and Watson, *Modern Analysis*, 2nd Ed., pp. 283, 333.

[17] Cf. R. A. Fisher, *loc. cit.*, I.

H. Hotelling, *Proc. National Academy of Sciences*, Vol. XI (1925), pp. 657–662.

[18] Cf. S. S. Wilks, *loc. cit.*, p. 482.

The characteristic function of the simultaneous distribution law of θ and θ_0 is immediately derivable from (5.14) by replacing t_0 by Nt_0 and t_r by $N_r t$ $(r = 1, 2, \cdots, k)$. There results

$$\varphi(t, t_0) = (1 - it_0)^{-(k-1)/2}(1 - it_0 - it)^{-(N-k)/2}. \tag{6.11}$$

By a procedure similar to that already used we find that the simultaneous distribution law of θ and θ_0 is given by

$$P(\theta, \theta_0) = \frac{\theta^{(N-k-2)/2}(\theta_0 - \theta)^{(k-3)/2}e^{-\theta_0}}{\Gamma(N-k)/2\,\Gamma(k-1)/2}. \tag{6.12}$$

By applying to (6.12) the transformation $\theta = \theta_0 h^2$, $\theta_0 = \theta_0$ and integrating out the value of θ_0, we find for the distribution law of h^2

$$D(h^2) = \frac{\Gamma(N-1)/2}{\Gamma(N-k)/2\,\Gamma(k-1)/2}(h^2)^{(N-k-2)/2}(1 - h^2)^{(k-3)/2}. \tag{6.13}$$

From (6.12) and (6.10) it may be shown that the following estimates of variance all have the same expected value[19]

$$\left\{\begin{aligned} &\frac{N_1V_1 + N_2V_2 + \cdots + N_kV_k}{N-k} \\ &\frac{NV_0}{N-1} \\ &\frac{N_1W_1 + N_2W_2 + \cdots + N_kW_k}{k-1} \end{aligned}\right. \tag{6.14}$$

7. Distribution of variances. Let

$$\left\{\begin{aligned} \theta_r &= N_rV_rB/D \qquad (r = 1, 2, \cdots, k) \\ \theta_0 &= NV_0B/D \\ \theta &= (B/D)(N_1V_1 + N_2V_2 + \cdots + N_kV_k) \end{aligned}\right. \tag{7.1}$$

where the right members of (7.1) are defined as in section 5. It is evident that the characteristic function of the simultaneous distribution law of θ, θ_0, θ_r, $(r = 1, 2, \cdots, k-1)$ is derivable from (5.14) by replacing t_0 by Nt_0, t_r by $N_r(t_r + t)$, $(r = 1, 2, \cdots, k-1)$ and t_k by $N_k t$. Thus

$$\begin{aligned} \varphi(t, t_0, t_1, \cdots, t_{k-1}) &= (1 - it_0)^{-(k-1)/2} \\ &\qquad (1 - it_0 - it)^{-(N_k-1)/2}\prod^{k-1}(1 - it_0 - it_\alpha - it)^{(1-N_\alpha)/2}. \end{aligned} \tag{7.2}$$

[19] Cf. J. Neyman and E. S. Pearson, II. *Biometrika*, Vol. 20A (1928), pp. 273–274. S. Kullback, *Annals of Mathematical Statistics*, Vol. 6 (1935), pp. 76–77.

By proceeding as in section 5 we arrive at the result that the simultaneous distribution law of θ, θ_0, θ_r, $(r = 1, 2, \cdots, k-1)$ is given by

$$
(7.3)\qquad P(\theta, \theta_0, \theta_r) = \frac{e^{-\theta_0}(\theta_0 - \theta)^{(k-3)/2}(\theta - \theta_1 - \theta_2 - \cdots - \theta_{k-1})^{(N_k-3)/2}}{\Gamma(k-1)/2\,\Gamma(N_k-1)/2} \prod_{\alpha=1}^{k-1} \frac{\theta_\alpha^{(N_\alpha-3)/2}}{\Gamma(N_\alpha-1)/2}
$$

where $\theta_0 \geqq \theta$, $\theta \geqq \theta_1 + \theta_2 + \cdots + \theta_{k-1}$.

By integrating out the variable θ_0 from (7.3) we have for the simultaneous distribution law of θ, θ_r, $(r = 1, 2, \cdots, k-1)$

$$
(7.4)\qquad D(\theta, \theta_r) = \frac{e^{-\theta}(\theta - \theta_1 - \theta_2 - \cdots - \theta_{k-1})^{(N_k-3)/2}}{\Gamma(N_k-1)/2} \prod_{\alpha=1}^{k-1} \frac{\theta_\alpha^{(N_\alpha-3)/2}}{\Gamma(N_\alpha-1)/2}
$$

A procedure similar to that used to derive (5.19) yields for the simultaneous distribution law of

$$
(7.5)\qquad \psi_r = \theta_r/\theta \qquad (r = 1, 2, \cdots, k-1)
$$

$$
P(\psi_1, \psi_2, \cdots, \psi_{k-1}) = \frac{\Gamma(N-k)/2}{\Gamma(N_k-1)/2}(1 - \psi_1 - \psi_2 - \cdots - \psi_{k-1})^{(N_k-3)/2} \prod_{\alpha=1}^{k-1} \frac{\psi_\alpha^{(N_\alpha-3)/2}}{\Gamma(N_\alpha-1)/2}
$$

where the limits of variation in (7.6) are[20]

$$
(7.7)\qquad \begin{cases} 0 \leqq \psi_1 \leqq 1 \\ 0 \leqq \psi_r \leqq 1 - \psi_1 - \psi_2 - \cdots - \psi_{r-1}, & (r = 2, \cdots, k-1). \end{cases}
$$

In a manner similar to the derivation of (6.6) we find the distribution law of $h_\alpha^2 = \psi_\alpha$, $(\alpha = 1, 2, \cdots, k-1)$, $h_k^2 = 1 - \psi_1 - \psi_2 - \cdots - \psi_{k-1}$ to be

$$
(7.8)\qquad D(h_\alpha^2) = \frac{\Gamma(N-k)/2}{\Gamma(N_\alpha-1)/2\,\Gamma(N-k-N_\alpha+1)/2} (h_\alpha^2)^{(N_\alpha-3)/2}(1-h_\alpha^2)^{(N-k-N_\alpha-1)/2}, \qquad (\alpha = 1, 2, \cdots, k).
$$

From the distribution law in (7.3) we readily obtain that the characteristic function of the distribution law of $\gamma_\alpha^2 = \log(\theta_\alpha/(\theta_0 - \theta))$ is given by

$$
(7.9)\qquad \varphi(t) = \frac{\Gamma(N_\alpha - 1 + 2it)/2\,\Gamma(k-1-2it)/2}{\Gamma(N_\alpha-1)/2\,\Gamma(k-1)/2} \qquad (\alpha = 1, 2, \cdots, k).
$$

[20] Cf. J. Neyman and E. S. Pearson, *loc. cit.*, 1.

We thus have that the distribution law of γ_α^2 is given by

$$(7.10)\quad P(\gamma_\alpha^2) = \frac{1}{\Gamma(N_\alpha - 1)/2\,\Gamma(k-1)/2} \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-it\gamma_\alpha^2}\,\Gamma(N_\alpha - 1 + 2it)/2\,\Gamma(k - 1 - 2it)/2\,dt.$$

The integral in (7.10) is known,[21] and there results

$$(7.11)\quad P(\gamma_\alpha^2) = \frac{\Gamma(N_\alpha + k - 2)/2}{\Gamma(N_\alpha - 1)/2\,\Gamma(k-1)/2}\, e^{\gamma_\alpha^2 (N_\alpha - 1)/2} \left(1 + e^{\gamma_\alpha^2}\right)^{-(N_\alpha + k - 2)/2}$$

If we set $e^{\gamma_\alpha^2} = \theta_\alpha/(\theta_0 - \theta) = \lambda_\alpha^2$ we have for the distribution of λ_α^2

$$(7.12)\quad D(\lambda_\alpha^2) = \frac{\Gamma(N_\alpha + k - 2)/2}{\Gamma(N_\alpha - 1)/2\,\Gamma(k-1)/2}\,(\lambda_\alpha^2)^{(N_\alpha - 3)/2}\,(1 + \lambda_\alpha^2)^{-(N_\alpha + k - 2)/2}$$

An extension of the procedure used to obtain (7.9) yields as the characteristic function of the simultaneous distribution of $\gamma_1^2, \gamma_2^2, \cdots, \gamma_k^2$

$$(7.13)\quad \varphi(t_1, t_2, \cdots, t_k) = \frac{\Gamma(k - 1 - 2it_1 - 2it_2 - \cdots - 2it_k)/2}{\Gamma(k-1)/2} \prod_{\alpha=1}^{k} \frac{\Gamma(N_\alpha - 1 + 2it_\alpha)/2}{\Gamma(N_\alpha - 1)/2}$$

Successive application of the method used to evaluate (7.10) yields as the simultaneous distribution law of the γ_α^2

$$(7.14)\quad P(\gamma_1^2, \gamma_2^2, \cdots, \gamma_k^2) = \frac{\Gamma(N-1)/2}{\Gamma(k-1)/2}\,(1 + e^{\gamma_1} + \cdots + e^{\gamma_k})^{-(N-1)/2} \prod_{\alpha=1}^{k} \frac{e^{\gamma_\alpha^2 (N_\alpha - 1)/2}}{\Gamma(N_\alpha - 1)/2}.$$

The simultaneous distribution of the λ_α^2 defined as in (7.12) is given by

$$(7.15)\quad D(\lambda_1^2, \lambda_2^2, \cdots, \lambda_k^2) = \frac{\Gamma(N-1)/2}{\Gamma(k-1)/2}\,(1 + \lambda_1^2 + \lambda_2^2 + \cdots + \lambda_k^2)^{-(N-1)/2} \prod_{\alpha=1}^{k} \frac{(\lambda_\alpha^2)^{(N_\alpha - 3)/2}}{\Gamma(N_\alpha - 1)/2}.$$

[21] Whittaker and Watson, *loc. cit.*, pp. 283, 383.

8. Conclusion. In this paper we have presented further instances of the applicability of the theory of characteristic functions to the distribution problem of statistics. In a subsequent paper the author hopes to illustrate the application of the results here developed to specific numerical problems.

WASHINGTON, D. C.

ON A CRITERION FOR THE REJECTION OF OBSERVATIONS AND THE DISTRIBUTION OF THE RATIO OF DEVIATION TO SAMPLE STANDARD DEVIATION

BY WILLIAM R. THOMPSON

Criteria for the rejection of outlying observations may be designed to reject a given fraction of all observations, or a proportion varying with the size of the sample. Irwin[1] has discussed several criteria based on sampling from a normal population which had been used previously, as well as one which he proposed. This is based on the principal of fixing the expectation of rejecting an observation from a sample independently of the aggregate number, N, of the sample. The criterion, λ, is $1/\sigma$ times the interval between successive observations in ascending order of magnitude, where σ is the standard deviation of the sampled population. In the same paper he gave, for different values of N, a table of $P_1(\lambda)$ and $P_2(\lambda)$, respectively probabilities of exceeding given values of λ for the first or second such interval from either end. In actual use, however, σ is estimated from the sample standard deviation, and we are left to decide whether observations in question are to be included or not in estimating the standard deviation as also whether or not to modify this by addition or subtraction of an estimate of its probable error. The object of the present communication is to develop a criterion free from defects of this nature, depending only on the assumption of random sampling from a normal universe. For this purpose we develop the distribution of τ defined by

(1)

where s is the sample standard deviation and δ is the deviation of an arbitrary observation of the sample from the sample mean. This leads to definite criteria, which are simple in application.

Accordingly, consider a sample $\{x_i\}$, $i = 1, \cdots, N$, to be drawn at random from a normal population of unknown mean and standard deviation, and that the order of enumeration is arbitrary. Then x_N is an arbitrary one of the elements or *observations*. Now, let

$$(2) \qquad \bar{x} = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i, \qquad s = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}}, \qquad \text{and}$$

$$(3) \qquad \delta \equiv x_N - \bar{x}.$$

Then we will prove that the distribution of $\tau \equiv \delta/s$ in repeated sampling with a fixed aggregate number, N, is given by substitution of

$$\sqrt{n}\cdot z = t = \sqrt{n}\cdot\tau/\sqrt{n+1-\tau^2}$$

in the z or t distribution of "Student" and R. A. Fisher,[2] where $n = N - 2$. To this end let $N > 2$, and let $n = N - 2$, and

$$(4) \qquad (n+1)\bar{x}_1 = \sum_{i=1}^{n+1} x_i, \quad \text{and} \quad S_1(x-\bar{x}_1)^2 = \sum_{i=1}^{n+1}(x_i - \bar{x}_1)^2.$$

Obviously, the $(n+1)\bar{x}_1 + x_N = N\cdot\bar{x}$, whence

$$(5) \qquad \bar{x} - \bar{x}_1 = \frac{x_N - \bar{x}}{n+1} = \frac{\delta}{n+1}, \qquad \text{whence} \qquad x_N - \bar{x}_1 = \frac{n+2}{n+1}\cdot\delta.$$

Furthermore, $N\cdot s^2 = S_1(x-\bar{x}_1)^2 + (n+1)(\bar{x}_1 - \bar{x})^2 + (x_N - \bar{x})^2$, whence

$$(6) \qquad N\cdot s^2 = S_1(x-\bar{x}_1)^2 + \frac{n+2}{n+1}\cdot\delta^2.$$

Now, considering the separate samples, $\{x_i\}$, $i = 1, \cdots, N-1$, and $\{x_N\}$, of aggregate number, $N - 1$ and 1, respectively; Fisher has shown[2] that if we set

$$(7) \qquad t = \frac{(x_N - \bar{x}_1)\cdot\sqrt{n}}{\sqrt{S_1(x-\bar{x}_1)^2}}\cdot\sqrt{\frac{n+1}{n+2}},$$

then, for $t_0 > 0$, the probability, p, that $t < t_0$ is

$$(8) \qquad p = \frac{1}{2} + \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\cdot\sqrt{n\cdot}}\int_0^{t_0}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}\cdot dt,$$

and $P = 2(1-p)$ is the probability that $|t| > t_0$.

Now, (5) and (6) in (7) give

$$(9) \qquad t = \frac{\frac{n+2}{n+1}\cdot\delta}{\sqrt{(n+2)\left(s^2 - \frac{\delta^2}{n+1}\right)}}\cdot\sqrt{\frac{n(n+1)}{n+2}} = \frac{\tau\cdot\sqrt{n}}{\sqrt{n+1-}}$$

whence

$$(10) \qquad \tau = t\sqrt{\frac{n+1}{n+t^2}}, \quad \text{or} \quad \frac{\tau}{\sqrt{n+1}} = \sin\theta, \qquad \tan\theta = \frac{t}{\sqrt{n}} = z.$$

Accordingly, P is the probability that $|\tau| > \tau_0 \equiv t_0\sqrt{\dfrac{n+1}{n+t_0^2}}$.

Thus, if we want to determine τ_0 so that by rejecting all observations deviating from the sample mean by more than $s\cdot\tau_0$ we shall have an average relative

frequency of rejections per sample which is fixed, say ϕ; then we need only to set $P = \phi/N$. This follows at once from the hypothesis as x is a random element of the random sample of N elements drawn from the same normal universe (of unknown mean and standard deviation). The criterion of rejection, $s \cdot \tau_0$, is uniquely determined from the sample standard deviation and

TABLE I

N	τ for given ϕ			t for given ϕ			n
	$\phi = 0.2$	0.1	0.05	0 2	0 1	0.05	
3	1.40646	1.41228	1.41373	9.51	19.08	38.19	1
4	1.6454	1.6887	1.7103	4.30	6.20	8.84	2
5	1.791	1.869	1.917	3.48	4.54	5 84	3
6	1.895	1.997	2.067	3.19	3.97	4 84	4
7	1.973	2.093	2.182	3.04	3.68	4.38	5
8	2.041	2.170	2.274	2.97	3.51	4.12	6
9	2.099	2.237	2.348	2 93	3.42	3 94	7
10	2.144	2.295	2.413	2.89	3.36	3 83	8
11	2.190	2.343	2.472	2.88	3 31	3 76	9
12	2.229	2.388	2.521	2.87	3 28	3 70	10
13	2.262	2.425	2.567	2 86	3 25	3 66	11
14	2.296	2.463	2 598	2 86	3 24	3 60	12
15	2.325	2.497	2 636	2 86	3 23	3 58	13
16	2.357	2.522	2 670	2 87	3 21	3.56	14
17	2.382	2 553	2.699	2 87	3 21	3 54	15
18	2.404	2 576	2 733	2 87	3 20	3 54	16
19	2.429	2 601	2 759	2 88	3 20	3 53	17
20	2.448	2.625	2 783	2 88	3 20	3 52	18
21	2.471	2 647	2 800	2 89	3 20	3 50	19
22	2 487	2.661	2 819	2 89	3 19	3 49	20
32	2 636	2 819	2 985	2 944	3 216	3.479	30
42	2.737	2 925	3.093	2 991	3 248	3.489	40
102	3.047	3 233	3 407	3 182	3 397	3 603	100
202	3.266	3.448	3.621	3 347	3 546	3.736	200
502	3.528	3 704	3.872	3 569	3.752	3.927	500
1002	3 714	3 881	4 047	3 737	3 908	4 078	1000

$P = \phi/N.$

Note: τ is computed to 0.5 unit in the last place given from the given t which is believed correct to 1 unit in the last place.

number of elements, N, for any prescribed ϕ. Dropping the subscript, critical values of τ are given in *Table I* (together with corresponding values of t) for $\phi = 0.2$, 0.1, and 0.05 and values of $n \equiv N - 2$ which should be sufficient for most practical purposes. The normal deviate (for unit standard deviation and the same P) lies between these values and is approached by τ and t (in the

tabulated range of ϕ) from opposite sides as n increases, the approximation to τ being the closer of the two. Accordingly Sheppard's tables may be used with good approximation for $n > 1000$, with $\phi/N = P$, the probability of exceeding numerically the given deviate. They may be used to advantage also in interpolation between $n = 100, 1000$ by means of differences at the tabulated points.

A crude rejection system where we reject an observation if it deviate from the mean of all others by more than a *fixed constant* times the standard deviation of such a difference in terms of σ as estimated from the variance of these others by $\tilde{\sigma} = \sqrt{\dfrac{S_1 (x - \bar{x}_1)^2}{N - 2}}$ amounts to taking a fixed value of t as criterion. The intention is usually to fix the probability (P) of rejection of observations rather than the expectation of rejections per sample (ϕ); and this, of course, is the expected approximate result for *large samples.* For small samples, however, say $4 < N < 32$, by rejection of observations deviating thus by more than $3 \cdot \tilde{\sigma} \cdot \sqrt{\dfrac{N}{N - 1}}$, it appears from (7) and *Table I* that approximately ϕ would be fixed rather than P.

The τ-criterion not only affords a precise extension of such a rejection system, but also a reduction of the actual process of application to a minimum, with one noteworthy exception for the case, $N = 3$. Here we may use as criterion with identical effect the ratio, $\dfrac{d_2}{d_1}$; where $x_1 \leqq x_2 \leqq x_3$, $d_2 = x_3 - x_2$, $d_1 = x_2 - x_1$, and $d_2 \geqq d_1$. This order can always be adopted for the test, and it is readily verified that

$$\frac{d_2}{d_1} = \frac{\sqrt{3} \cdot t - 1}{2}, \tag{11}$$

whence for $\phi = 0.2, 0.1$, and 0.05, respectively we have $\dfrac{d_2}{d_1} \cong 7.74, 16.0$, and 32.6. Thus, for $N = 3$, we may take merely the ratio of the greater to the other numerical deviation from the median observation as criterion.

Section 2

Although not required in connection with the rejection criterion developed above, there is a simple generalization of τ with a closely related distribution which may be valuable in somewhat different circumstances. Consider the same situation as given above, except that $\{x_i\}$ is divided into two subsets, where $i = 1, \cdots, N - k$, and $i = N - k + 1, \cdots, N$, respectively; giving two random samples of aggregate number, $N - k$ and k. Let the means of these be $\bar{x}_1$ and $\bar{x}_2$, respectively; and s and $\bar{x}$ be as before. Then in general let

$$\delta \equiv \bar{x}_2 - \bar{x} \quad \text{and} \quad \tau \equiv \frac{\delta}{s}. \tag{12}$$

TABLE II

$\tau_{(P,N,1)}$

N	P = 0.9	0 8	0 7	0 6	0 5	0 4	P = 0 3	0 2	0.1	0 05	0 02	0.01	N
3	221	437	643	832	1 000	1 144	1 260	1 3450	1 3968	1 4099	1 41352	1 414039	3
4	173	347	520	693	866	1 039	1 212	1 386	1 559	1 6080	1 6974	1 7147	4
5	158	316	476	639	808	983	1 170	1 374	1 611	1.757	1.869	1 9175	5
6	.149	300	453	612	.777	.952	1 143	1 360	1 631	1.814	1 973	2.0509	6
7	144	290	440	.594	.757	932	1 125	1 349	1 640	1 848	2 040	2.142	7
8	.141	284	431	583	744	918	1 111	1 340	1 644	1 870	2 087	2.207	8
9	139	280	425	.575	734	907	1 102	1 334	1 647	1 885	2 121	2 256	9
10	.137	.276	420	569	727	899	1 094	1 328	1 648	1 895	2 146	2 294	10
11	136	.274	416	.564	.721	893	1 088	1 324	1 648	1 904	2 166	2 324	11
12	.135	272	413	.560	.717	.888	1 083	1 320	1.649	1 910	2 183	2 348	12
13	.134	270	411	557	713	884	1 080	1 317	1 649	1 915	2 196	2 368	13
14	134	269	408	554	710	881	1 076	1 314	1 649	1 919	2 207	2 385	14
15	133	268	407	552	.707	878	1 073	1 312	1 649	1 923	2 216	2 399	15
16	.133	267	405	550	705	875	1 071	1 310	1 649	1 926	2 224	2 411	16
17	.132	.266	404	548	703	.873	1 069	1 309	1 649	1 928	2 231	2 422	17
18	.132	.265	403	547	701	871	1 067	1 307	1.649	1 931	2 237	2 432	18
19	131	264	402	546	699	869	1 065	1 305	1 649	1.932	2 242	2 440	19
20	131	264	401	544	698	.868	1 063	1 304	1 649	1 934	2 247	2 447	20
21	130	263	400	543	697	.867	1 062	1 303	1 649	1 936	2 251	2 454	21
22	.130	.263	399	542	696	865	1 061	1 302	1 649	1 937	2 255	2 460	22
23	.130	262	398	.541	695	864	1 059	1 301	1 649	1 938	2 259	2 465	23
24	.130	.262	.398	541	694	863	1 058	1 300	1 649	1 940	2 262	2 470	24
25	.130	.261	397	540	693	862	1 057	1 299	1 649	1 941	2 264	2 475	25
26	.130	.261	397	539	692	861	1 056	1 299	1 648	1.942	2 267	2 479	26
27	129	.261	.397	538	691	860	1 056	1 298	1 648	1 942	2 269	2 483	27
28	129	.261	396	.538	691	860	1 055	1 297	1 648	1 943	2 272	2 487	28
29	.129	260	.396	537	690	859	1 054	1 297	1 648	1 944	2 274	2 490	29
30	.129	260	395	537	690	859	1 054	1 296	1 648	1 944	2 275	2 493	30
31	.129	.260	.395	536	689	858	1 054	1 296	1 648	1 945	2 277	2 495	31
32	129	.260	394	536	.689	858	1 053	1 295	1 648	1 945	2 279	2 498	32
∞	.12566	.25335	38532	.52440	.67449	84162	1 03643	1 28155	1 64485	1 95996	2 32634	2 57582	∞

Note: $\tau_{(P,N,k)} \equiv \sqrt{\dfrac{N-k}{k(N-1)}} \cdot \tau_{(P,N,1)}$

Further, let $n_1 + 1 = N - k$, $n_2 + 1 = k$, $S_1(x - \bar{x}_1)^2$ be the sum of squared deviations in the first sub-sample and similarly $S_2(x - \bar{x}_2)^2$ be that for the second. Then Fisher has shown[2] that the generalized

$$(13) \qquad t = \frac{(\bar{x}_2 - \bar{x}_1)\sqrt{n_1 + n_2}}{\sqrt{S_1(x - \bar{x}_1)^2 + S_2(x - \bar{x}_2)^2}} \sqrt{\frac{(n_1 + 1)(n_2 + 1)}{n_1 + n_2 + 2}}$$

is distributed as before for $n = n_1 + n_2$. Obviously,

$$N \cdot \bar{x} = (n_1 + 1)\bar{x}_1 + (n_2 + 1)\bar{x}_2 ,$$

whence

$$\delta \equiv \frac{(n_1 + 1)(\bar{x}_2 - \bar{x}_1)}{N} \equiv \frac{(n_1 + 1)(\bar{x} - \bar{x}_1)}{n_2 + 1}, \tag{14}$$

and

$$\begin{aligned} S_1(x - \bar{x}_1)^2 + S_2(x - \bar{x}_2)^2 &= N \cdot s^2 - (n_1 + 1)(\bar{x}_1 - \bar{x})^2 - (n_2 + 1)(\bar{x}_2 - \bar{x})^2 \\ &= N\left(s^2 - \frac{n_2 + 1}{n_1 + 1} \cdot \delta^2\right), \end{aligned} \tag{15}$$

whence

$$t = \tau \sqrt{\frac{n \cdot k}{n + 2 - k - k \cdot \tau^2}}, \qquad \text{where } n = N - 2; \tag{16}$$

i.e., $t = \sqrt{n} \cdot \tan \theta$, $\sqrt{n + 2 - k} \cdot \sin \theta = \sqrt{k} \cdot \tau$.

In connection with analysis of variance where the total sample may be divided into several subsets of observations, the generalized τ may be used, accordingly, to indicate in a simple manner which (if any) of the means of subsets differ significantly from the general mean where the equivalent t-test is applicable.

In general let $\tau_{(P, N, k)} \geqq 0$ be a number such that P is the probability that $/\tau/ > \tau_{(P, N, k)}$; where, as above, N is the total number of observations in the whole sample, k is the number of these in the subsample and τ is defined by (12). Then by (16), obviously,

$$\tau_{(P, N, k)} \equiv \sqrt{\frac{N - k}{k(N - 1)}} \cdot \tau_{(P, N, 1)}. \tag{17}$$

In *Table II* are given values of $\tau_{(P, N, 1)}$ for a range of values of the arguments, N and P. The critical values of τ in *Table I* are simply values of this function for $P = \phi/N$ where ϕ is taken as parameter, i.e., $\tau_{(\phi/N, N, 1)}$.

Rider[3] has given an interesting review of rejection criteria previously proposed.

BIBLIOGRAPHY

1. IRWIN, J. O., *Biometrika, 17* (1925), pp. 100–128; *17* (1925), pp. 238–250.
2. FISHER, R. A., *Metron, 5*, (1925), No. 3, pp. 90–104, 109–112.
   "Student," *Metron, 5*, (1925), No. 3, pp. 105–108, 113–120.
3. RIDER, PAUL R., St. Louis, Washington University Studies, new series, Science and Technology, No. 8 (1933), 23 pp.

YALE UNIVERSITY.

ON CERTAIN COEFFICIENTS USED IN MATHEMATICAL STATISTICS

By Everett H. Larguier, S.J.

I. Introduction

(1.1) We have studied here certain coefficients arising in interpolation, numerical differentiation and integration formulas in order to establish explicit expansions for these coefficients in the form of a finite summation. Ordinarily they are obtained by means of recursion relations, which necessarily demand the building up of a complete table in order to find the desired set of coefficients. By using the methods described in this paper, we are able to calculate any desired set independent of the ones which precede it in the table. In the literature we find two other expansions of the *difference quotients of zero*, one by Jeffery[1] and one by Boole.[2] Our expansion for the *differential quotients of zero* is the same as one obtained by Jeffery,[3] however the proof is more elementary and simple.

The Bernoulli numbers also find a wide range of application in many finite integration formulas, and hence our attention was drawn to the discussion of certain coefficients which occur in the study of these functions.[4] As in the cases mentioned above these coefficients are likewise ordinarily obtained by recursion formulas, but by our expansions they may be obtained directly.

II. Difference Quotients of Zero

(2.1) It is our purpose here to show that this difference quotient of zero, $\Delta^m 0^n$, may be expressed by the following summation:

$$\Delta^m 0^n = m! \sum \left(\frac{m}{m-1}\right)^{a_{m-1}} \left(\frac{m-1}{m-2}\right)^{a_{m-2}} \cdots \left(\frac{3}{2}\right)^{a_2} \left(\frac{2}{1}\right)^{a_1} \tag{1}$$

where $a_1, a_2, \cdots, a_{m-1} = 0, 1, 2, \cdots, n-m$ and $a_1 \geqq a_2 \geqq \cdots \geqq a_{m-1} \geqq 0$. Obviously the number of terms in the summation is the number of combinations of $n - m + 1$ things taken $m - 1$ together where repetitions are allowed.

(2.2) By means of the recursion relation[5]

$$\Delta^m 0^n = m\, \Delta^m 0^{n-1} + m\, \Delta^{m-1} 0^{n-1} \tag{2}$$

[1] Henry M. Jeffery, "On a method of expressing the combinations and homogeneous products of numbers and their powers by means of differences of nothing." *Quarterly Journal of Pure and Applied Mathematics*, vol. 4 (1861), pp. 364 ff.

[2] George Boole, *A Treatise on the Calculus of Finite Differences*, (Stechert, N. Y.), p. 20.

[3] Loc. cit.

[4] Steffensen, *Interpolation* (Williams & Wilkins, Baltimore), p. 125.

[5] L. M. Milne-Thompson, *Calculus of Finite Differences*, (Macmillan), p. 36, sec. 2.53, (2).

we are able to build up a table of values. By substitution it can be shown that (1) satisfies the values of this table except when $m = 0, 1$ and for $m > n$, for then the summation becomes meaningless. We therefore define the summation to have the value 0 for $m = 0, n > 0$ and for $m > n$, and the value 1 for $m = 1$. We exhibit one substitution below. When $m = 3$ and $n = 4$,

$$\Delta^3 0^4 = 3!\left\{\left(\frac{3}{2}\right)^0\left(\frac{2}{1}\right)^0 + \left(\frac{3}{2}\right)^0\left(\frac{2}{1}\right)^1 + \left(\frac{3}{2}\right)^1\left(\frac{2}{1}\right)^1\right\} = 36 .$$

(2.3) Taking (2), we proceed by repeated application of the recursion formula and finally we have

$$\Delta^m 0^n = m^{n-m}\,\Delta^m 0^m + \sum_{d=m}^{n-1} m^{n-d}\,\Delta^{m-1} 0^d ,$$

which since $\Delta^m 0^m = m!$,[6] becomes

$$\Delta^m 0^m = m^{n-m}\, m! + \sum_{d=m}^{n-1} m^{n-d}\,\Delta^{m-1} 0^d . \tag{3}$$

We will now prove (1). Proceeding by induction we assume (1) true for $m - 1$. Hence from (3) we have

$$\Delta^m 0^n = m^{n-m}\, m! + \sum_{d=m}^{n-1} m^{n-d}\,(m-1)! \sum \left(\frac{m-1}{m-2}\right)^{a_{m-2}} \cdots \left(\frac{3}{2}\right)^{a_2}\left(\frac{2}{1}\right)^{a_1} ,$$

where $a_1, a_2, \cdots, a_{m-2} = 0, 1, 2, \cdots, d - m + 2$ and $a_1 \geqq a_2 \geqq \cdots \geqq a_{m-2} \geqq 0$. This becomes

$$\Delta^m 0^n = m^{n-m}\, m! + m! \sum_{d=m}^{n-1} m^{n-d-1} \sum \left(\frac{m-1}{m-2}\right)^{a_{m-2}} \cdots \left(\frac{3}{2}\right)^{a_2}\left(\frac{2}{1}\right)^{a_1} . \tag{4}$$

Using the symbol $\Sigma\Sigma$ for the double summation of (4), we may write

$$\begin{aligned}\sum\sum = \sum_{d=m}^{n-1} m^{n-d-1} &\left\{\left(\frac{m-1}{m-2}\right)^0 \cdots \left(\frac{3}{2}\right)^0\left(\frac{2}{1}\right)^0 + \left(\frac{m-1}{m-2}\right)^0 \cdots \left(\frac{3}{2}\right)^0\left(\frac{2}{1}\right)^1\right.\\ &+ \left(\frac{m-1}{m-2}\right)^0 \cdots \left(\frac{3}{2}\right)^1\left(\frac{2}{1}\right)^1 + \cdots\\ &+ \left(\frac{m-1}{m-2}\right)^{d-m}\left(\frac{m-2}{m-3}\right)^{d-m+1} \cdots \left(\frac{3}{2}\right)^{d-m+1}\left(\frac{2}{1}\right)^{d-m+1}\\ &\left.+ \left(\frac{m-1}{m-2}\right)^{d-m+1} \cdots \left(\frac{2}{1}\right)^{d-m+1}\right\}\\ = \sum_{d=m}^{n-1} &\left\{\left(\frac{m}{m-1}\right)^{n-d-1} \cdots \left(\frac{3}{2}\right)^{n-d-1}\left(\frac{2}{1}\right)^{n-d-1}\right.\end{aligned}$$

[6] Milne-Thompson, loc. cit.

$$+\left(\frac{m}{m-1}\right)^{n-d-1}\cdots\left(\frac{3}{2}\right)^{n-d-1}\left(\frac{2}{1}\right)^{n-d}+$$

$$+\left(\frac{m}{m-1}\right)^{n-d-1}\left(\frac{m-1}{m-2}\right)^{n-m-1}\left(\frac{m-2}{m-3}\right)^{n-m}\cdots\left(\frac{2}{1}\right)^{n}$$

$$+\left(\frac{m}{m-1}\right)^{n-d-1}\left(\frac{m-1}{m-2}\right)^{n-m}\cdots\left(\frac{2}{1}\right)^{n-}$$

Now, $m^{n-m} = \left(\frac{m}{m-1}\right)^{n-m}\left(\frac{m-1}{m-2}\right)^{n-m}\cdots\left(\frac{3}{2}\right)^{n-m}\left(\frac{2}{1}\right)^{n-m}$, and also d varies from m to $n-1$. Hence by including m^{n-m} under the summation we are able to replace the double summation by a single one and have

$$\Delta^m 0^n = m!\sum\left(\frac{m}{m-1}\right)^{a_{m-1}}\left(\frac{m-1}{m-2}\right)^{a_{m-2}}\cdots\left(\frac{3}{2}\right)^{a_2}\left(\frac{2}{1}\right)^{a_1}$$

where $a_1, a_2, \cdots, a_{m-1} = 0, 1, 2, \cdots, n-m$ and $a_1 \geqq a_2 \geqq \cdots \geqq a_{m-1} \geqq 0$. Hence (1) is proved.[7]

III. Differential Quotients of Zero

(3.1) In Markoff's formula for numerical differentiation we meet coefficients of the type $D^m 0^{(n)}$. We will show here that this differential quotient of zero may be expressed by the following finite sum:

$$D^m 0^{(n)} = (-1)^{n-m}\, m!\sum (p_1 p_2 \cdots p_{n-m}) \tag{5}$$

where $p_1 > p_2 > \cdots > p_{n-m} > 0$ take on values from $1, 2, \cdots, n-1$. Obviously the number of terms in the expansion will be the same as the number of combinations of $n-1$ things taken $n-m$ together without repetitions.

(3.2) By means of the recursion formula[8]

$$D^m 0^{(n)} = (1-n)\, D^m 0^{(n-1)} + m\, D^{m-1} 0^{(n-1)} \tag{6}$$

we are able to build up a table of values. By substitution it can easily be shown that (5) satisfies the values of the table when $n > m > 0$. For the other values the summation is meaningless, hence we define it to have the value 1 for $m = n > 0$; and the value 0 for $m > n$ and $m = 0$. When $m = 2$ and $n = 4$, we have

$$D^2 0^{(4)} = (-1)^{4-2}\, 2!\,\{(3\cdot 2) + (3\cdot 1) + (2\cdot 1)\} = 22,$$

which is the same value as found by (6).

[7] Our expansion may be shown to be equal to that of Jeffery's cited in the introduction, which is $\Delta^m 0^{m+n} = m!\,\xi^m 0^{m+n}$, where $\xi^m 0^{m+n}$ expresses the sum of all the homogeneous products of n dimensions which can be formed by the first m natural numbers and their powers. The proof of Jeffery's expansion involves the use of complicated symbolic operators, while our proof uses elementary notions only.

[8] Steffensen, op. cit., p. 57, 58, (12) and (14).

(3.3) Returning to (6), we obtain by its repeated application:

$$D^m0^{(n)} = (-1)^{n-m}(n-1)^{(n-m)}D^m0^{(m)} + m\sum_{a=0}^{n-m-1}(-1)^a(n-1)^{(a)}D^{m-1}0^{(n-a-1)}$$

or, since $D^m0^{(m)} = m!$,

$$D^m0^{(n)} = (-1)^{n-m}(n-1)^{(n-m)}m! + m\sum_{a=0}^{n-m-1}(-1)^a(n-1)^{(a)}D^{m-1}0^{(n-a-1)} \quad (7)$$

In proving (5), we proceed by induction, assuming (5) true for $m-1$; hence by (7) we have

$$\begin{aligned} D^m0^{(n)} &= (-1)^{n-m}(n-1)^{(n-m)}m! \\ &\qquad + m!\sum_{a=0}^{n-m-1}(-1)^{n-m}(n-1)^{(a)}\sum(p_1p_2\cdots p_{n-m-a}) \end{aligned} \quad (8)$$

where $p_1 > p_2 > \cdots > p_{n-m-a} > 0$ take the values $1, 2, \cdots, n-a-2$. Expanding the double sum of (8) we have

$$\begin{aligned} \sum\sum = \sum_{p_1=1}^{n-2}(p_1\cdots p_{n-m}) &+ \sum_{p_1=1}^{n-3}(n-1)(p_1\cdots p_{n-m-1}) \\ &+ \sum_{p_1=1}^{n-4}(n-1)(n-2)(p_1\cdots p_{n-m-2}) \\ &+ \cdots + \sum_{p_1=1}^{m-1}(n-1)(n-2)\cdots(m+1)(p_1) \end{aligned} \quad (9)$$

in which $p_1 > p_2 > \cdots > p_s > 0$ always holds, where

$$s = n-m, n-m-1, \cdots, 2, 1$$

in turn.

Upon inspection, it is evident that (9) contains all the terms of (5) with the exception of $(n-1)(n-2)\cdots(m+1)m$. Hence, since by definition $(n-1)^{(n-m)} = (n-1)\cdots(m+1)m$, we may include the first term on the right-hand side of (8) under the summation and then we have proved (5).[9]

IV. The Coefficient $G_n^{(r)}$

(4.1) In discussing the Bernoulli numbers and the Bernoulli polynomials, Steffensen[10] makes use of the relation:

$$B_{2r}(x) = (-1)^r\sum_{n=0}^{r}G_n^{(r)}z^{r-n} \quad (10)$$

[9] Jeffery's expansion referred to in the introduction is $D^m0^{(n)} = \zeta^m0^{(n)}$, where $\frac{(-1)^{n-m}\zeta^m0^{(n)}}{m!}$ expresses the sum of the combinations of the first $n-1$ natural numbers taken $n-m$ together. The remarks made above under article 2.3 concerning symbolic operators also apply here *mutatis mutandis*.

[10] Op. cit., p. 125, (24); cf. also Jacobi's theorem. *Journal für reine und angewandte Mathematik* (Crelle's Journal), vol. 12, pp. 268–269.

where $z = x - x^2$. We wish here to show that the coefficient $G_n^{(r)}$, ordinarily found by means of recursion formulas, may be obtained from the following summation:

$$G_n^{(r)} = (2r)^{(2n)} \sum_{N_n=3}^{r-n+1} [N_n] \sum_{N_{n-1}=3}^{N_n+1} [N_{n-1}] \cdots \sum_{N_1=3}^{N_2+1} [N_1] \tag{11}$$

where $[N] = (N)^{(2)}/(2N)^{(4)}$. Obviously the summation has no meaning for $n = 0$, nor for $r < n + 2$. Therefore it will be necessary to make definitions or devise other schemes for meeting this difficulty.

Steffensen[11] shows that

$$G_0^{(r)} = 1 \quad \text{for} \quad r \geqq 0; \qquad G_{r-1}^{(r)} = 0 \quad \text{for} \quad r > 1; \tag{12}$$

and likewise he gives the following recursion relation:

$$(2r - 2n)^{(2)} G_n^{(r)} = (2r)^{(2)} G_n^{(r-1)} + (r - n + 1)^{(2)} G_{n-1}^{(r)}. \tag{13}$$

In accordance with (12), we define the sum of (11) to be equal to 1 for $n = 0$, and to be equal to 0 for $n = r - 1$, when $r > 1$. By means of the recursion formula (13), Steffensen[12] gives a table of values of $G_n^{(r)}$, which (11) may be easily shown to satisfy. From this table we have the value $G_3^{(6)} = 10$. Using this as an example of the expansion, we have by (11):

$$\begin{aligned} G_3^{(6)} &= (12)^{(6)} \sum_{N_3=3} [N_3] \sum_{N_2=3}^{N_3+1} [N_2] \sum_{N_1=3}^{N_2+1} [N_1] \\ &= (12)^{(6)} \langle [3]\{[4]([5] + [4] + [3]) + [3]([4] + [3])\} \\ &\qquad + [4]\{[5]([6] + [5] + [4] + [3]) + [4]([5] + [4] + [3]) + [3]([4] + [3])\}\rangle \\ &= 10. \end{aligned}$$

(4.2) Before proving the general case, we will prove by induction that

$$G_1^{(r)} = (2r)^{(2)} \sum_{N_1=3}^{r} [N_1] \tag{14}$$

Assuming (14) true for $r - 1$, we have by (12) and (13)

$$G_1^{(r)} = (2r)^{(2)} \sum_{N_1=3}^{r-1} [N_1] + (2r)^{(2)} [r] = (2r)^{(2)} \sum_{N_1=3}^{r} [N_1].$$

Hence (14) is valid.

(4.3) We shall prove (11) with respect to r. By repeated application of (13), we have

[11] Op. cit., p. 125.
[12] Op. cit., p. 126.

$$\begin{aligned}
G_n^{(r)} &= \{(2r)^{(2)}/(2r-2n)^{(2)}\}G_n^{(r-1)} + \{(2r)^{(2)}(r-n+1)^{(2)}/(2r-2n+2)^{(4)}\}G_{n-1}^{(r-1)} \\
&\quad + \{(2r)^{(2)}(r-n+1)^{(2)}(r-n+2)^{(2)}/(2r-2n+4)^{(6)}\}G_{n-2}^{(r-1)} + \cdots \\
&\quad + \{(2r)^{(2)}(r-n+1)^{(2)} \cdots (r-1)^{(2)}/(2r-2)^{(2n)}\}G_1^{(r-1)} \\
&\quad + \{(r-n+1)^{(2)} \cdots (r)^{(2)}/(2r-2)^{(2n)}\}G_0^{(r)} \\
&= (2r)^{(2n)} \sum_{N_n=3}^{r-n} [N_n] \cdots \sum_{N_1=3}^{N_2+1} [N_1] \\
&\quad + (2r)^{(2n)}[r-n+1] \sum_{N_{n-1}=3}^{r-n+1} [N_{n-1}] \cdots \sum_{N_1=3}^{N_2+1} [N_1] \\
&\quad + (2r)^{(2n)}[r-n+1][r-n+2] \sum_{N_{n-2}=3}^{r-n+2} [N_{n-2}] \cdots \sum_{N_1=3}^{N_2+1} [N_1] + \cdots \\
&\quad + (2r)^{(2n)}[r-n+1][r-n+2] \cdots [r-1] \sum_{N_1=3}^{r-1} [N_1] \\
&\quad + (2r)^{(2n)}[r-n+1] \cdots [r].
\end{aligned}$$

It is evident from inspection that this is nothing but an expanded form of (11), hence (11) is proved with respect to r.

(4.4) Proceeding in the same way as above to prove induction with respect to n, we have again by repeated application of (13)

$$\begin{aligned}
G_n^{(r)} &= \{(r-n+1)^{(2)}/(2r-2n)^{(2)}\}G_{n-1}^{(r)} + \{(2r)^{(2)}(r-n)^{(2)}/(2r-2n)^{(4)}\}G_{n-1}^{(r-1)} \\
&\quad + \{(2r)^{(4)}(r-n-1)^{(2)}/(2r-2n)^{(6)}\}G_{n-1}^{(r-2)} \\
&\quad + \cdots + \{(2r)^{(2r-2n-4)}(3)^{(2)}/(2r-2n)^{(2r-2n-2)}\}G_{n-1}^{(n+2)} \\
&= (2r)^{(2n)}[r-n+1] \sum_{N_{n-1}=3}^{r-n+2} [N_{n-1}] \cdots \sum_{N_1=3}^{N_2+1} [N_1] \\
&\quad + (2r)^{(2n)}[r-n] \sum_{N_{n-1}=3}^{r-n+1} [N_{n-1}] \cdots \sum_{N_1=3}^{N_2+1} [N_1] \\
&\quad + \cdots + (2r)^{(2n)}[4] \sum_{N_{n-1}=3}^{5} [N_{n-1}] \cdots \sum_{N_1=3}^{N_2+1} [N_1] \\
&\quad + (2r)^{(2n)}[3] \sum_{N_{n-1}=3}^{4} [N_{n-1}] \cdots \sum_{N_1=3}^{N_2+1} [N_1].
\end{aligned}$$

From this latter equation, (11) follows immediately and therefore the proof is complete.

(4.5) Bernoulli numbers may be expressed in terms of this coefficient $G_n^{(r)}$, as is shown by Steffensen,[13] in the following way

$$B_{2r} = (-1)^r G_r^{(r)} \tag{15}$$

[13] Op. cit., p. 125, (27).

which we shall express in terms of (11). However as (11) is meaningless for $n = r$, we obtain the relation

$$(2r + 2)^{(2)} G_r^{(r)} = -(2)^{(2)} G_{r-1}^{(r+1)} \quad \text{for} \quad r > 0, \tag{16}$$

which follows immediately from (12) and (13), and thereby obviate this difficulty. Hence, by (11), (15) and (16), we can write

$$B_{2r} = \{(-1)^{(r+1)}(2r)!/(4)^{(2)}\} \sum_{N_{r-1}=3}^{3} [N_{r-1}] \sum_{N_{r-2}=3}^{N_{r-1}+1} [N_{r-2}] \cdots \sum_{N_1=3}^{N_2+1} [N_1] \tag{17}$$

We note here that the definitions of the summation, given in 4.1, likewise hold.

SAINT LOUIS UNIVERSITY
SAINT LOUIS, MISSOURI

NOTICE OF THE ORGANIZATION OF THE INSTITUTE OF MATHEMATICAL STATISTICS

For sometime there has been a feeling that the theory of statistics would be advanced in the United States by the formation of an organization of those persons especially interested in the mathematical aspects of the subject. As a consequence, a meeting of interested persons was arranged for September 12, 1935, at Ann Arbor, Michigan. At the meeting, it was decided to form an organization to be known as the Institute of Mathematical Statistics. A constitution and by-laws were adopted and the following officers elected to serve until December 31st, 1936: President, H. L. Rietz; Vice-president, W. A. Shewhart; Secretary-Treasurer, A. T. Craig. A resolution, instructing the officers to investigate the feasibility of the affiliation of the Institute with the American Mathematical Society or with the American Statistical Association, was adopted.

The constitution provides that membership in the Institute shall consist of Members, Fellows, Honorary Members, and Sustaining Members. A committee on membership will establish qualifications requisite for the different grades of membership. The annual dues of members and fellows are five dollars a year and these include a year's subscription to the official journal, the Annals of Mathematical Statistics.

The next meeting of the Institute will be held in St. Louis, Missouri, in December of this year in connection with the meetings of the American Association for the Advancement of Science, the American Mathematical Society, and other organizations.

Forms for application for membership in the Institute may be had by writing the Secretary-Treasurer at the University of Iowa, Iowa City, Iowa.

1935 DIRECTORY OF SUBSCRIBERS TO THE ANNALS OF MATHEMATICAL STATISTICS

INDIVIDUALS

ACERBONI, DR. ARGENTINO V., Banfield, Larroque 232, U. T. 94, Argentine.
ARMSTRONG, CHARLES M., JR., 1338 Dean Street, Schenectady, N. Y.
ANDERSON, ROBERT, 1104 New Federal Building, St. Paul, Minn.
AROIAN, LEO. A., Colorado State College, Fort Collins, Colo.
BACHELOR, ROBERT W., 1437 Bancroft Way, Berkeley, Calif.
BAILEY, W. B., The Travelers Insurance Company, Hartford, Conn.
BATEN, PROF. W. D., University of Michigan, Ann Arbor, Mich.
BEALE, DR. FRANK S., 316 West Packer Avenue, Bethlehem, Penn.
BEEN, RICHARD, United States Bureau of Agricultural Economics, Washington, D. C.
BLACK, LOREN T., R. F. D. No. 3, Peru, Indiana.
BLACKADAR, W. L., Assistant Actuary, Equitable Life Assurance Society, New York City.
BLACKBURN, PROF. RAYMOND F., University of Pittsburg, Pittsburg, Penn.
BORGWARD, PROF. F. WM., College of Business Administration, Syracuse University, Syracuse, N. Y.
BROWN, PROF. THEODORE H., School of Business Administration, Harvard University, Boston, Mass.
BOSCH, J., Barcelona RDA Universidad, Barcelona, Spain.
BRUMBAUGH, PROF. M. A., Bureau of Business and Social Research, University of Buffalo, Buffalo, N. Y.
BUSHEY, PROF. J. HOBART, Hunter College, New York City.
CARMICHAEL, JOHN, 2 Aldridge Road, Villas, London W. 11, England.
CARVER, PROF. H. C., University of Michigan, Ann Arbor, Mich.
CHAPMAN, DR. DWIGHT W., Emerson Hall, Cambridge, Mass.
CHAPMAN, ROY A., Southern Forest Experimental Station, New Orleans, La.
CLIFFE, F. B., General Electric Company, Schenectady, N. Y.
COGGINS, PAUL P., Statistician, New Jersey Bell Telephone Company, Newark, N. J.
COOVER, PROF. J. E., Department of Psychology, Stanford University, Calif.
CRAIG, PROF. ALAN T., University of Iowa, Iowa City, Iowa.
CRAIG, PROF. C. C., University of Michigan, Ann Arbor, Mich.
CRATHORNE, PROF. A. R., University of Illinois, Urbana, Ill.
CURETON, E. E., Associate Professor of Education, Alabama Polytechnic Institute, Auburn, Alabama.
DADE, E. B., 116 W. Administration Building, Lawrence, Kansas.
DODD, PROF. EDWARD L., University of Texas, Austin, Texas.
DORWEILER, PAUL, Actuarial Department, Aetna Life Insurance Company, Hartford, Conn.
DUNLAP, JACK W., Fordham University, New York City.
ELSTON, JAMES S., Travelers Insurance Company, Hartford, Conn.
EVANS, HERBERT P., Mathematics Department, University of Wisconsin, Madison, Wisc.
FERTIG, DR. JOHN W., Assistant Biometrician, State Hospital, Worcester, Mass.
FISCHER, DR. CARL H., Mathematics Department, Wayne University, Detroit, Mich.
FLETCHER, R. C., Director, Bureau of Social Research, Federation of Social Agencies, Pittsburg, Penn.
FRECHEVILLE, G., Windmill Hill Farm, Steeple Claydon, Bucks, England.

FRITZE, C. E., Fredagatan 2, Stockholm, Sweden.
GLOVER, PROF. J. W., University of Michigan, Ann Arbor, Mich.
GREVILLE, DR. THOS. N. E., Actuarial Department, Acacia Mutual Life Insurance Company, Washington, D. C.
GRIFFIN, HAROLD D., Nebraska State Teachers College, Wayne, Nebr.
HAMILTON, PROF. THOS. R., Texas Agricultural and Mechanical College, College Station, Texas.
HAMMOND, H. PIERSON, Travelers Insurance Company, Hartford, Conn.
HARCOURT, C. J., Assistant Vice President, New York Telephone Company, New York City.
HARMON, G. E., M. D., Board of Health, Chicago, Ill.
HENDERSON, ROBERT, Vice President and Actuary, Equitable Life Assurance Society, New York City.
HENDRICKS, WALTER A., Junior Biologist, United States Animal Husbandry Experimental Station, Beltsville, Maryland.
HENRY, MALCOLM H., Mathematics Department, Michigan State College, East Lansing, Mich.
HILDEBRANDT, PROF. E. H. C., State Teachers College, Upper Montclair, N. J.
HILL, M. A., JR., Mathematics Department, University of North Carolina, Chapel Hill, N. C.
HOTELLING, PROF. HAROLD, Department of Economics, Columbia University, New York City.
HUNTINGTON, PROF. EARL V., 48 Highland Street, Cambridge, Mass.
INGRAHAM, PROF. MARK H., University of Wisconsin, Madison, Wisc.
KEFFER, RALPH, Associate Actuary, Aetna Life Insurance Company, Hartford, Conn.
KELLEY, PROF. TRUMAN L., Harvard University, Cambridge, Mass.
KELLOGG, LESTER S., College of Commerce and Administration, Ohio State University, Columbus, Ohio.
KOZELKA, RICHARD L., School of Business, University of Minnesota, Minneapolis, Minn.
KULLBACK, SOLOMON, 5311 8th Street, N.W., Washington, D. C.
LARSEN, MISS OLGA, Florida State College for Women, Tallahassee, Fla.
LARSON, HAROLD D., Mathematics Department, University of New Mexico, Albuquerque, N. M.
LEAVENS, DICKSON H., Harvard Business School, Soldiers Field, Boston, Mass.
MACKINNON, JOSEPH C., Registrar, Massachusetts Institute of Technology, Cambridge, Mass.
MAHALANOBIS, PROF. P. C., 210 Cornwallis Street, Calcutta, India.
MALZBERG, DR. BENJ., Senior Statistician, New York State Department of Mental Hygiene, Albany, N. Y.
MCEWEN, GEO. F., Schripps Institute of Oceanography, University of California, La Jolla, Calif.
MINER, DR. JOHN RICE, School of Hygiene and Public Health, Johns Hopkins University, Baltimore, Md.
MOLINA, EDWARD C., Switching Theory Engineer, Bell Telephone Laboratories, New York City.
MOWBRAY, PROF. ALBERT H., University of California, Berkeley, Calif.
MUDGETT, PROF. BRUCE D., School of Business, University of Minnesota, Minneapolis, Minn.
NEIFELD, DR. NORRIS R., 883 Brooklyn Avenue, Brooklyn, N. Y.
NESS, PROF. MARIE M., 127 Medical Sciences Building, University of Minnesota, Minneapolis, Minn.
NYSWANDER, PROF. JAMES A., University of Michigan, Ann Arbor, Mich.
OLIVERO, JOSE BOSCH, Apartado no. 991, Barcelona, Spain.

O'TOOLE, DR. A. L., College of St. Catherine, St. Paul, Minn.
OWEN, DR. F. V., United States Department of Agriculture, Salt Lake City, Utah.
PAYNE, DR. CHARLES K., Washington Square College, New York University, New York City.
PRETORIUS, S. J., Prima, Stellenbosch, South Africa.
REGAN, DR. FRANCIS, St. Louis University, St. Louis, Mo.
RICE, DR. J. NELSON, 3326 13th Street, N.E., Washington, D. C.
RIDER, DR. PAUL R., University College, Gower Street, London, W. C. 1, England.
RIETZ, PROF. H. L., University of Iowa, Iowa City, Iowa.
RIETZ, DR. WM., Scientific Research Division, United States Public Health Service, Washington, D. C.
ROBB, RICHARD A., 27 Moor Road, Eaglesham, Renfrewshire, Scotland.
ROBERTS, JOHN L., 28 Boody Street, Brunswick, Maine.
ROHM, JOHN T., Actuary, American Life Insurance Company, Detroit, Mich.
RORTY, COL. MALCOLM C., Old Spout Farm, Lusby, Calvert County, Maryland.
ROSS, DR. FRANK A., 405 Fayerweather Hall, Columbia University, New York City.
ROWELL, MISS DOROTHY C., 91 Bishop Street, New Haven, Conn.
RULON, PHILIP J., 4 Emerson Hall, Harvard University, Cambridge, Mass.
SAFFIAN, MISS SADIE, Room 205, 330 S. 9th Street, Philadelphia, Penn.
SCHULTZ, PROF. HENRY, Department of Economics, University of Chicago, Chicago, Ill.
SCHUMACHER, FRANCIS X., United States Forest Service, Washington, D. C.
SEELEY, BURTON D., Resettlement Administration, Washington, D. C.
SHELTON, W. ARTHUR, United States Department of Agriculture, Bureau of Public Roads, Washington, D. C.
SHEWHART, DR. WALTER A., Bell Telephone Laboratories, New York City.
SKONBERG, CARL M., Builders and Manufacturers Mutual Casualty Company, Chicago, Ill.
SMITH, HARTLEY LE HURAY, 500 Kent Avenue, Brooklyn, N. Y.
SOUTO, DR. JOSE BARRAL, Cordoba, 1535, Buenos Aires, Argentine.
SPRINGER, WILLIAM M., Bristol-Myers Company, Research Department, Hillside, N. J.
STEPHAN, FREDERICK F., 722 Woodward Building, Washington, D. C.
TAITEL, MARTIN, 130 Summers Drive, Alexandria, Va.
TAO, L. K., Institute of Social Sciences, 3 Wen Tsin Chieh, Hsi An Men, Peiping, China.
TOOPS, PROF. HERBERT A., Department of Psychology, Ohio State University, Columbus, Ohio.
TOPIC, F., 11 Narodni, Prague, Czechoslovakia.
TREANOR, PROF. GLEN R., School of Business Administration, University of Minnesota, Minneapolis, Minn.
TRELOAR, ALAN E., Biometric Laboratory, University of Minnesota, Minneapolis, Minn.
TRZASKA, EVERT MICHALSKI, Krakowskie Przidiescie 13, Ksiegarnia w Warszawie, Poland.
TRYON, FREDERICK G., Mineral Economist, Bureau of Mines, Washington, D. C.
WAITE, DR. WARREN C., Director of Agricultural Economics, University Farm, St. Paul, Minn.
WALKER, PROF. HELEN M., Teachers College, Columbia University, New York City.
WEIDA, PROF. FRANK M., George Washington University, Washington, D. C.
WETMORE, RAYMOND E., Bureau of Business Research, Metropolitan Life Insurance Company, New York City.
WICKSELL, PROF. S. D., Statistiska Institutionen vid Lunds Universitet, Lund, Sweden.
WINSOR, CHARLES P., 123 Holden Green, Cambridge, Mass.
WRIGHT, DR. SEWALL, Department of Zoology, University of Chicago, Chicago, Ill.
YNTEMA, DR. THEODORE O., University of Chicago, Chicago, Ill.
ZOCH, RICHMOND T., United States Weather Bureau, Washington, D. C.

LIBRARIES

Agnes Scott Library, Decatur, Ga.
Aktiebolaget Nordiska Bokhandeln, Drottninggatan 7, Stockholm, Sweden.
Antioch College Library, Yellow Springs, Ohio.
Armour Institute of Technology Library, Chicago, Ill.
Bell Telephone Laboratories, Technical Library, 463 West Street, New York City.
Bibli. de la Faculdad de Derecho, Universidad Literaria, Valencia, Spain.
Bombay University, Bombay, India.
Brown University Library, Providence, R. I.
Buchhandlung des Oesterreichischen Bundesverlages, Schwarzenbergstr. 5, Wien 1, Austria.
Bureau of Statistics, Ministry of Industries, Nanking, China.
Carnegie Institute of Technology, Schenley Park, Pittsburg, Penn.
California Institute of Technology, Mathematics Library, Pasadena, Calif.
Columbia University, New York, N. Y.
Connecticut College, New London, Conn.
Dartmouth College, Hanover, N. H.
Dominion Bureau of Statistics, Sussex Street, Ottawa, Canada.
Fordham University Graduate School, Woolworth Building, New York, N. Y.
Fundament Bibli., Lesotechnitsch Akademii, Leningrad 18, U. S. S. R.
Fund. IB-KA Sredne-Asiaisk, Gisud. Uniwersiteta, Potschtowi J Jastschik No. 47, Taschkent, U. S. S. R.
Giannini Foundation Library, University of California, Berkeley, Calif.
Goucher College, Baltimore, Md.
Guthrie and Co., Ltd., Java Street, Kuala Lumpur, Federated Malay States.
Hamilton Smith College, University of New Hampshire, Durham, N. H.
Harvard College Library, Cambridge, Mass.
Hawaiian Sugar Planters Association, Experimental Station, Honolulu, Hawaii.
Hirschwaldsche Buchhandlung, Unter den Linden 68, Berlin, N.W. 7, Germany.
Imperial Institute of Agricultural Research, Pusa, Bihar, India.
Indiana University, Bloomington, Ind.
Institute of Social Sciences, Academia Sinica, Peiping, China.
Iowa State College, Ames, Iowa.
Jackson, Son and Company, 73 West George Street, Glasgow, Scotland.
Lehigh University Library, Bethlehem, Penn.
Lehman Corp., 1 So. William St., New York, N. Y.
Library of Congress, Washington, D. C.
Libreria Victoriano Suarez, Biblioteca de Estadistica, Madrid, Spain.
Maharajah's College, Mysore, India.
Mandelle Memorial Library, Kalamazoo College, Kalamazoo, Mich.
Mathamatisches Inst. der Univ., Bunsenstr. 3–5, Göttingen, Germany.
Michigan State College, East Lansing, Mich.
Montana State College, Bozeman, Montana.
Nankai University, Tientsin, China.
National Resources Com., 2 San Yuan Hong, Nanking, Kiangsu, China.
National Research Council, Sussex Street, Ottawa, Canada.
Nautschn Bib-Ka Zunchu, B. Wusowskij per 2, Moskwa, U. S. S. R.
New York Public Library, 5th Avenue and 42nd Street, New York City.
New York State College of Agriculture, Agricultural Economics and Farm Management Department, Ithaca, N. Y.
North Dakota State College, Fargo, North Dakota.
Northwestern University, Evanston, Ill.

Ohio State University, Columbus, Ohio.
Peiping Union Medical College, Peiping, China.
Porter Library, Kansas State Teachers College, Pittsburg, Kansas.
Post Biblioteki, Akademii Nauk U.S.S.R, Bolsaja Koluzskaja, 24, Moskwa, U.S.S.R.
Presidency College, Statistical Laboratory, Calcutta, India.
Purdue University, Lafayette, Indiana.
Rensselaer Polytechnic Institute, Troy, N. Y.
Rockefeller Foundation, 49 West 49th Street, New York City.
Royal Statistical Society, 9 Adelphi Terrace, London, W.C. 2, England.
Smith College, Northampton, Mass.
Southern Forest Experimental Station, 400 Union Building, New Orleans, La.
Stadt Bibliothek, Schöne Aussicht 2, Frankfurt A. M., Germany.
Stanford University, Stanford, Calif.
Statens Skogsförsöksanstalt Biblioteket, Experimentalfältet, Sweden.
Tsing Hua University, Peiping, China.
United States Department of Agriculture, Washington, D. C.
Universita Commerciale Luigi Bocconi, Via Solferino, Milano, Italy.
University of Arkansas, Engineering Library, Fayetteville, Arkansas.
University of Buffalo, Edmund Hayes Hall, Buffalo, N. Y.
University of California at Los Angeles, Mathematics Library, Los Angeles, Calif.
University of California at Los Angeles, University Library, Los Angeles, Calif.
University of California, Berkeley, Calif.
University of Chicago, Harper Memorial Library, Chicago, Ill.
University of Cincinnati, Burnet Woods Park, Cincinnati, Ohio.
University of Illinois, Urbana, Ill.
University of Iowa, Iowa City, Iowa.
University of Kansas, Periodical Department, Lawrence, Kansas.
University of Kentucky, Lexington, Ky.
University of London, Department of Applied Statistics, University College, London England.
University of Manitoba, Fort Garry Site, Winnipeg, Canada.
University of Maryland, College Park, Maryland.
University of Michigan, Ann Arbor, Michigan.
University of Minnesota, Minneapolis, Minn.
University of Nebraska, Lincoln, Nebr.
University of Oklahoma, Norman, Okla.
University of Oregon, Eugene, Oregon.
University of Pennsylvania, Philadelphia, Penn.
University of Pittsburg, 310 State Hall, Pittsburg, Penn.
University of Pretoria, Pretoria, South Africa.
University of Rochester, Rochester, N. Y.
University of Texas, Austin, Texas.
University of Toronto, Toronto 5, Canada.
University of Washington, Seattle, Washington.
Washington University, St. Louis, Mo.
Wesleyan University, Middletown, Conn.
Western Reserve University, Cleveland, Ohio.
Williams College, Williamstown, Mass.
Yale University, New Haven, Conn.
Zentr. Nautschnaja s/ch Biblioteka, Orlikow per 1/11, Moskwa, U. S. S. R.
Zentralblatt-Bureau, Julius Springer, Linkstrasse 23–24, Berlin, Germany.

Date of issue.	Date of issue.	Date of issue.